

# THIS WEEK

## EDITORIALS

**WORLD VIEW** It is time to scale up supply of unproven Ebola drugs **p.233**

**SALMON RUN** Exactly how many farmed fish really escape to the sea? **p.234**



**ROBOTS** Lego-inspired bricks that make flexible friends **p.234**

## What goes up

*Federal restrictions on the use of drones by US researchers threaten an increasingly productive tool. The scientific community must speak out while there is a chance to change matters.*

When US regulators ordered the journalism programme at the University of Missouri in Columbia to stop using camera-carrying remotely piloted aircraft last year, researchers around the country watched in alarm. The drones had been flown over private property, with the consent of the landowner. They had remained below 120 metres' altitude to avoid interfering with larger aircraft. Most thought that such flights would be legal.

They are not, according to the US Federal Aviation Administration (FAA). It is well known that the agency has largely banned commercial drone flights, pending the development of regulations to ensure their safety. But, as we report on page 239, many scientists did not realize that the agency considers research and education at private universities to be 'commercial' activities. It is an unfortunate and distorted definition that threatens research programmes in a wide swathe of disciplines. Scientists must speak out to make the FAA aware of these impacts.

There is much to say. The FAA has an unfeasibly narrow definition of those eligible to apply for special permission to fly unmanned aircraft. The agency has applied its historical division between government- and civil-operated aircraft to universities, creating a nonsensical distinction between public universities that receive a substantial amount of government funding and private universities that do not. Researchers at public universities are eligible to apply for an exemption to the commercial flight ban; researchers at private universities are not.

It is clear that the FAA has a difficult job. Technological advances are making drones increasingly appealing for everything from police work to package delivery. The agency must forge regulations that will ensure the safety and propriety of the machines and how they are used. A smattering of drone accidents — including the crash of a tourist's drone into a famous hot spring in Yellowstone National Park on 2 August — underscore the importance of these regulations. The agency should also take care not to hamper the burgeoning field's development.

Researchers must make their needs heard amid the clamour of lobbyists from industrial-drone manufacturers and aeroplane-pilot unions. Yet many researchers remain unaware that their work is threatened. Some continue to fly their machines in blissful ignorance of the FAA's rules. Others knowingly flout the guidance.

The community needs to spread the word — both to its own members and to the FAA — about the threats to research if drone use remains restricted. On 23 June, the FAA announced guidance intended to clarify its stance on drones, and outlining the distinctions that concern researchers. That document is open for public comment until 23 September, providing a clear opportunity to voice concerns to the agency.

The FAA is hard at work developing its regulations for drones, and intends to release an initial draft before the end of the year. That draft will also be open for public comment, but scientists need not wait until then to offer the agency their input. It is important to guide the discussion before it is too late to change its course. ■

## Finding the root

*The NIH is right to investigate whether bias makes grant awards unfair.*

A prominent 2011 paper in *Science* found that white researchers receive grants from the US National Institutes of Health (NIH) at nearly twice the rate that African American researchers do (D. K. Ginther *et al. Science* 333, 1015–1019; 2011). Although some of the disparity could be explained by differences in education, institution and publication record, the sheer magnitude of the result seemed to suggest that something more insidious was at play.

The idea that scientists who volunteer time and energy to review NIH grants could be biased against qualified minority researchers is a tough pill to swallow. The NIH is to be commended for not sweeping this possibility under the rug; it has turned to the scientific method to investigate the suggestion. Over the next three years, researchers will strip names and other identifying information from grant applications,

text-mine proposals for subtle clues to an applicant's race that might subconsciously tip off reviewers, and study reviewer critiques to see whether they hold evidence of prejudice (see page 243).

It is a topic that the NIH will need to broach delicately. Few academics consciously hold any such inclinations, and fewer still would deliberately allow them to affect their grant evaluations. Some are likely to bristle at what might be seen as an accusation of racism, and the NIH plans to conduct at least some of its studies of grant reviews without the reviewers' knowledge or consent.

But better for the NIH to offend a few people than to make snap judgements and institute blunt policies to address the problem. Fixes such as increasing scholarships and training for minority groups would no doubt be a good thing, but they could be an unhelpful use of money if they do not address the root cause of the disparity. And policies such as grant-allocation quotas could come at the expense of other researchers.

The NIH says that it will be guided by the data produced. Interventions could include training programmes on bias for reviewers, or using peer reviewers from different demographic groups, such as early-career scientists. If the agency does find evidence of bias, fixing it will be a difficult task. Unconscious bias, wherever it resides, is a difficult thing to turn off, even for the most educated and progressive of people. ■

KATE BRADY



## Scale up the supply of experimental Ebola drugs

*Estimates of the probable impact of the outbreak show that existing stocks of potentially useful medicines are insufficient, says Oliver Brady.*

With the worst-ever outbreak of Ebola raging in West Africa, a World Health Organization (WHO) committee last week concluded that it is ethical to use unproven drugs and vaccines to try to combat the disease, as long as doctors obtain patients' informed consent. There are no medicines currently approved for routine use against Ebola, either to treat infected people or to protect those they come into contact with, so we are in uncharted territory. Two logical and immediate questions are: what investigational drugs and vaccines are available, and what volume of each would be required?

At the front line, options for therapy and post-exposure treatment include passive immunization with monoclonal or polyclonal antibodies, and antiviral agents. For broader protection, several vaccines have been tested on non-human primates.

With the backing of the WHO, policy-makers and funders are now trying to decide which of these options to accelerate into active service. They need good estimates of how many of these drugs and vaccines to manufacture and distribute to control an Ebola outbreak.

Together with colleagues, I have been trying to provide such estimates.

We have separated the people who require help into four categories. Most urgently, there are those who have already become infected with Ebola virus and people close to them, such as family members. Next are the medical and support staff who treat patients, and those who handle the corpses. At less immediate risk but still important to protect are essential non-medical staff in the region of the outbreak, such as humanitarian-aid workers and people who provide key local services. A case can be made that protection should also be offered to key domestic government workers and others providing essential logistical support. Finally, we have already seen isolated cases of Ebola spread far from its West African source by travellers, and policy-makers should consider protection for these imported cases.

The scientific literature holds some information about probable levels of exposure in these groups (see [go.nature.com/1le6ua](http://go.nature.com/1le6ua)). This provides the best available evidence base for political and private funding decisions on the volume of drugs or vaccines that would be required.

To make this information available, my colleagues and I have constructed a spreadsheet that calculates the total number of people that might require treatment for a given outbreak (see [go.nature.com/vv98gv](http://go.nature.com/vv98gv)). This value is customizable depending on factors such as which of the above categories are to be targeted.

The intention is not to provide exact numbers of doses required, but rather to scope potential demand for a number of realistic scenarios.

This demand is likely to be higher than many

people realize. For example, our analysis suggests that, even under a conservative scenario, up to 30,000 people would have so far required treatment or prophylaxis in the current outbreak — substantially more than in any previous outbreak. The difference reflects the scale of the current emergency, which has made the jump from rural to urban areas. The WHO warned last week that reported numbers of cases and deaths “vastly underestimate” the size of the problem.

To estimate the demand for therapeutic or prophylactic agents more accurately, more-detailed data on patient contact rates and health-care-worker exposure must be collected or made available by the relevant organizations. These factors are likely to change as the Ebola epidemic spreads, treatment centres become available and people are quarantined.

Our estimates may need to be increased if, with the transition from rural to urban environments, infected people are coming into contact with more people. Under such conditions, tracking a person's contacts for the full recommended 21 days after exposure to the disease becomes logistically challenging, and it may be necessary to refine which contacts are defined as epidemiologically significant. Policy-makers should consider the role of strategies such as mass vaccination and greater use of personal protective equipment.

Our analysis is crude and has very clear limitations. But it does demonstrate that for treatment and prevention interventions to be rolled out evenly and fairly, stocks must be scaled up substantially. It seems that supplies of the monoclonal-antibody therapy ZMapp are already

exhausted, and available stocks of many other investigational drugs are limited to treatment courses for tens or hundreds of people, rather than the required thousands or tens of thousands.

It is clear that the scale of the current outbreak presents a change in the development landscape for those invested in Ebola therapeutics. As well as the direct disease burden, the unfolding epidemic in West Africa has revealed the huge potential for indirect costs brought about by political destabilization and crippled health-care services.

The use of ZMapp has already raised issues of equity of access to potentially life-saving therapies. But as WHO assistant director-general Marie-Paule Kieny has said: “I don't think that there could be any fair distribution of something that exists in such a small quantity.”

The scale of the ongoing outbreak may tilt the politics and economics to speed the development of a drug or vaccine. But it also makes it difficult to scale up production and distribution. All involved must rise to meet the challenge. ■

**Oliver Brady** is an epidemiologist at the University of Oxford, UK.  
e-mail: [oliver.brady@zoo.ox.ac.uk](mailto:oliver.brady@zoo.ox.ac.uk)

UNDER A  
CONSERVATIVE  
SCENARIO, UP TO  
**30,000 PEOPLE**  
WOULD HAVE SO FAR  
REQUIRED  
**TREATMENT OR  
PROPHYLAXIS.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/urry3v](http://go.nature.com/urry3v)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## HIV

### Antibody–drug mix stops relapse

A combination of antibodies and multiple virus-activating drugs can keep HIV from resurging in infected mice, even after treatment ends.

During drug treatment, HIV enters a dormant state and stays hidden inside infected cells; afterwards, it bounces back. A team led by Michel Nussenzweig at the Rockefeller University in New York tested a combination of neutralizing antibodies and three drugs that activate dormant HIV so it is no longer hidden. After the treatment was given to HIV-infected mice, 57% were protected from virus resurgence, whereas no significant effects were seen in mice treated with antibodies plus only one inducer or with antibodies alone.

Such drug combinations could reduce the reservoir of HIV-infected immune cells, a key step towards curing the disease, the authors say.

**Cell** <http://doi.org/t7w> (2014)

## MATERIALS

### Soft machines made like Lego

Soft, stretchy, Lego-style bricks offer a way to make three-dimensional (3D) prototypes of elastic structures, according to researchers at Harvard University in Cambridge, Massachusetts.



‘Click-e-bricks’, which were developed by George Whitesides and his colleagues, can be used to build stretchy devices, such as hollow ones that expand when air is injected (**pictured**) or that have internal channels for liquid. The approach could be used to rapidly make prototypes of soft machines, such as soft robots, that move depending on changes in air pressure, current or light.

The team argues that click-e-bricks offer a faster alternative to 3D printing, which relies on hard acrylic

polymers that limit the composition and complexity of the final structure.

**Adv. Mater.** <http://doi.org/f2tdnq> (2014)

## ASTRONOMY

### Comets forge organic molecules

Astronomers have captured three-dimensional images of organic compounds streaming from two comets.

Comets contain some of the oldest materials in the Solar System. Using the

collected data on how many tagged animals were recaptured by fishermen. In some cases, immature fish were found near their release sites, but for adults, the recapture rate after 1–2 years was less than 0.1%.

Using this probability for recapturing escapees, the team estimates that as many as 1.5 million farmed salmon escape from farms in Norway each year — significantly more than the 413,000 escapees that are reported annually.

**ICES J. Mar. Sci.** <http://doi.org/t6t> (2014)



## OCEAN SCIENCES

### Farmed salmon swim to freedom

Vastly more salmon could be escaping from aquaculture farms (pictured) than is officially reported, say Ove Skilbrei and his colleagues at the Institute of Marine Research in Bergen, Norway.

Farmed salmon that escape could mate with wild populations and make them less fit for survival. The researchers tagged more than 90,000 farmed Atlantic salmon (*Salmo salar*) and released them along the Scandinavian coast in 2005. Over the next five years, the team

JEAN GAUMY/MAGNUM PHOTOS

STEPHEN A. MORIN, HARVARD UNIV.

compounds could have been important for kicking off the chemistry that led to life on Earth, the researchers say. *Astrophys. J. Lett.* 792, L2 (2014)

## MICROBIOLOGY

## How *Salmonella* bounces back

Two groups have shown how *Salmonella* bacteria can resist antibiotics.

Dirk Bumann of the University of Basel in Switzerland and his colleagues infected mice with modified *Salmonella* strains that glow green when they divide. They found varying rates of division in different tissues, and most of the bacteria that survived antibiotic treatment had a moderate growth rate.

In a separate study, M  d  ric Diard at the Swiss Federal Institute of Technology in Zurich and his co-workers found that whereas antibiotics kill off less-dangerous *Salmonella* mutants in the mouse gut cavity, the more-virulent strains escape by hiding inside the gut tissue. After antibiotic treatment ended, the more-virulent bacteria repopulated the gut.

The findings could point to new strategies for antibiotic treatment, the authors say. *Cell* 158, 722–733 (2014); *Curr. Biol.* <http://doi.org/t7z> (2014)

## ASTRONOMY

## Dusty visitors from interstellar space

Seven particles captured by NASA's Stardust spacecraft may be the first sample of dust from beyond the Solar System that has been brought back to Earth.

Andrew Westphal at the University of California, Berkeley, and his colleagues — with the help of 30,714 citizen scientists around the world — scanned more than 1 million images of tracks left by particles on Stardust's collectors. They identified seven candidates with a surprising variety of crystal structures and elemental compositions.

These particles could help to explain the origin and evolution of interstellar dust, which current astronomical observations can only guess at. *Science* 345, 786–791 (2014)

## CONSERVATION BIOLOGY

## Poaching leads to elephant decline

The illegal killing of elephants in Africa to supply the ivory trade has reached unsustainable rates.

George Wittemyer at Colorado State University in Fort Collins and his colleagues used data from elephant carcass surveys in 45 sites across Africa to model broader trends in elephant poaching on the continent. They found that levels of illegal killing peaked in 2011 at 8% of the global population, leading to the loss of roughly 40,000 elephants, or a 3% overall reduction in animal numbers that year.

Decreasing demand for ivory in China is key to conserving elephants, say the authors. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1403984111> (2014)

## VIROLOGY

## Secret to Ebola's success

The Ebola virus might elude immune responses by stopping a key protein in infected cells from activating defence genes.

Ebola, which kills up to 90% of people it infects, is known to disrupt the activity of interferon, a crucial antiviral protein. Gaya Amarasinghe at Washington University School of Medicine in St Louis, Missouri, and his colleagues found that an Ebola viral protein blocks the transport of an interferon-activated protein called STAT1 into the cell nucleus. STAT1 is needed in the nucleus to stimulate defence mechanisms.

The results suggest new drug targets in the ongoing fight against the virus. *Cell Host Microbe* 16, 187–200 (2014)

## SOCIAL SELECTION

Popular articles on social media

## Strong words over a 'Hobbit'

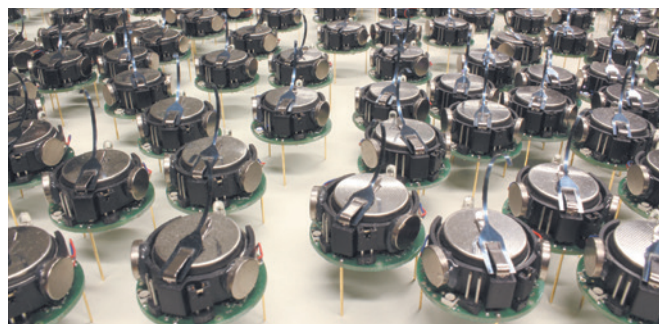
Two papers stirred up the palaeoanthropology world by suggesting that *Homo floresiensis* — a putative human relative discovered on the Indonesian island of Flores in 2003 — was instead an example of *Homo sapiens* with Down's syndrome. The theory, in the *Proceedings of the National Academy of Sciences*, was greeted with much scepticism. As part of a string of tweets, anthropologist Holly Dunsworth at the University of Rhode Island in Kingston said: "Conclusion [is] based seemingly on zilch." Co-author Robert Eckhardt, a geneticist at Pennsylvania State University, defended the diagnosis in a comment posted on a blog of the Natural History Museum in London, saying that his group and others have spent the past decade "trying to turn the 'Hobbit' circus into science".

*Proc. Natl Acad. Sci. USA* <http://doi.org/t66>; <http://doi.org/t65> (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/qizfui](http://go.nature.com/qizfui)



## ENGINEERING

## Robot swarms take shape

A thousand-strong army of coin-sized robots (pictured) can arrange itself into various configurations.

Michael Rubenstein and his co-workers at Harvard University in Cambridge, Massachusetts, programmed 1,024 robots with a simple set of rules and an image of a shape to be formed. Four 'seed robots' act as a point of origin for a coordinate system and send their coordinates to neighbours using infrared light. This information spreads through the group, allowing each robot to determine its relative location in the swarm.

The robot flock can form programmed shapes — such as the letter 'K' — in around 12 hours and is the largest yet to demonstrate collective behaviour, the authors say. *Science* 345, 795–799 (2014)

## CLARIFICATION

The Research Highlight 'Brain scans predict TV hits' (*Nature* 512, 8; 2014) notes that Jacek Dmochowski is at Stanford University; however, the research described was done at the City College of New York.

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## POLICY

### European data law

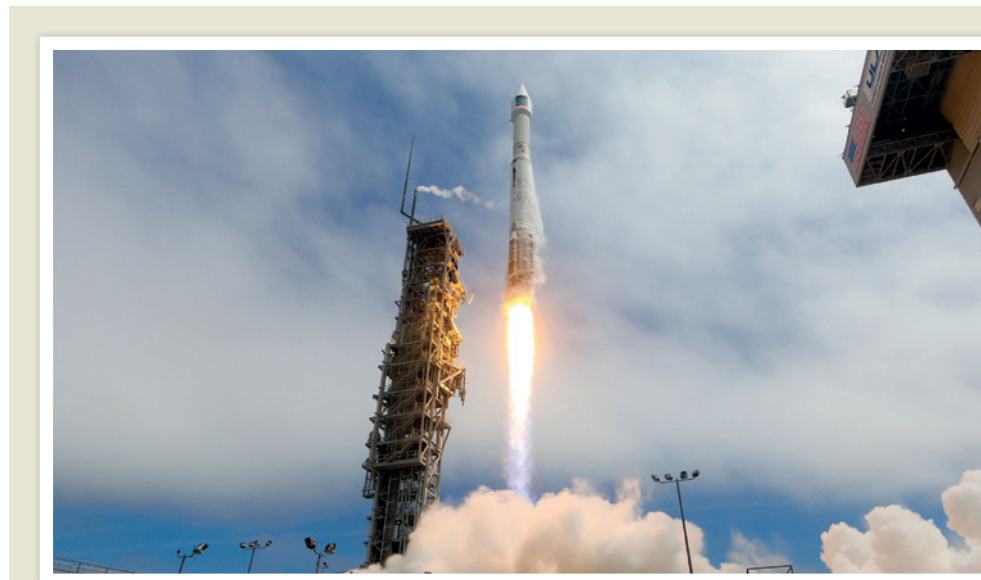
Proposed amendments to the law that regulates the use of personal data in the European Union will stifle research, warned leading research organizations in an open letter on 12 August (see [go.nature.com/jchze8](http://go.nature.com/jchze8)). The letter argues that the changes would make the law too heavy-handed in its requirements for individual consent for the use of personal data in research, and “would limit our ability to ensure European policy making is informed by the most robust evidence base”. Negotiations on the legislation, which is expected to be finalized by the end of the year, are reaching their final stages.

### Forest threat

Changes in Indonesia's government could put the country's progress in forest protection at risk, says an 18 August report from the Norwegian government. In 2010, Norway — the largest funder of the United Nations' Reducing Emissions from Deforestation and Forest Degradation (REDD) programme — pledged US\$1 billion to help protect Indonesia's forests. Indonesia recently overtook Brazil to have the world's highest rate of deforestation.

### Endangered species

The US Fish and Wildlife Service has dropped its bid to protect wolverines (*Gulo gulo*) under the Endangered Species Act. Wolverines inhabit snow-covered areas, including mountains; the agency last year proposed to list the species as being endangered by the effects of climate change on its habitat. On 12 August, the service announced that it was withdrawing the proposal, saying that “climate change



PATRICK H. CORKERY/LOCKHEED MARTIN CORPORATION/UNITED LAUNCH ALLIANCE

## Satellite brings Earth into focus

A commercial satellite that will produce the sharpest images of Earth yet was launched on 13 August by DigitalGlobe, a company in Longmont, Colorado. The WorldView-3 satellite will image the planet at a resolution of 31 centimetres per pixel. The company previously had to limit the images it sells to consumers to 50-centimetre resolution, until the US Department of Commerce relaxed its

rules in June to allow the sale of images with 25-centimetre resolution. The move means that Earth imaging will finally catch up with that of Mars — NASA's Mars Reconnaissance Orbiter has been capturing the red planet at a similar resolution to WorldView-3's since 2006. The improved imagery will enable more-precise mapping of farmland and forests, and will help mining and oil companies in exploration.

models are unable to reliably predict snowfall amounts and snow-cover persistence in wolverine denning locations”. Polar bears and two types of ice-dwelling seal are the only animals listed because of the impacts of climate change.

## EVENTS

### Ebola crisis

The Ebola epidemic in West Africa has worsened. As of 16 August, 1,229 deaths had been reported in Guinea, Nigeria, Liberia and Sierra Leone, but the World Health Organization (WHO) said that is probably an underestimate. On 16 August, a mob attacked a quarantine centre in Monrovia, Liberia. Meanwhile,

the Canadian government said that it would donate hundreds of doses of an experimental vaccine to the WHO for use in West Africa, and three infected Liberian health workers were given a different experimental treatment called ZMapp. The WHO concluded that it is ethical to use unproven drugs against Ebola, but warned against “unrealistic expectations”. See page 233 for more.

### Researcher suicide

A suicide note left by a leading Japanese stem-cell researcher says that a media onslaught around the retraction of two papers that he co-authored led him to take his own life on 5 August, a lawyer for

his family said in a press conference on 12 August. Yoshiki Sasai was a renowned scientist at the RIKEN Center for Developmental Biology in Kobe, Japan. He co-authored two ostensibly groundbreaking papers that were published in *Nature* in January, but retracted in July after problems with the data emerged. A RIKEN investigation committee had found Sasai to hold “grave responsibility” for poor oversight of the work, but concluded that he was innocent of involvement with the problematic data. It found the papers' lead author, Haruko Obokata, guilty of misconduct. See [go.nature.com/2godrn](http://go.nature.com/2godrn) for more.



## FACILITIES

## Research ship

Australia's purpose-built research vessel, *Investigator*, will leave a shipyard in Singapore on 25 August for its home port of Hobart in Tasmania. Officials from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) took possession of the Aus\$122-million (US\$113-million) ship on 4 August, months later than expected owing to construction delays. *Investigator* can accommodate up to 40 scientists and will venture farther south than its predecessor, *Southern Surveyor*. But budget cuts at CSIRO mean that the ship will operate for only 180 days a year, rather than the 300 days researchers had hoped for.

## AWARDS

## Fields medals

Mathematician Maryam Mirzakhani (pictured) became the first woman to win a Fields Medal, the highest honour in mathematics. The International Mathematical Union revealed the four 2014 medal winners on 12 August. Fields medals, which each come with a Can\$15,000 (US\$13,700) prize — are awarded every four years to researchers aged 40 or younger. Mirzakhani, who is at Stanford



University in California, won for her work on “the dynamics and geometry of Riemann surfaces and their moduli spaces”. The three other medallists are Manjul Bhargava of Princeton University in New Jersey, Martin Hairer of the University of Warwick, UK, and Artur Avila of the Institute of Mathematics of Jussieu in Paris. See [go.nature.com/sklvtd](http://go.nature.com/sklvtd) for more.

## RESEARCH

## Costs of misconduct

Scientific misconduct has wasted around US\$58 million of funding from the US National Institutes of Health (NIH) between 1992 and 2012, according to a study published on 14 August (A. M. Stern *et al.* *eLife* 3, e02956; 2014). Ferric Fang from the University of Washington, Seattle, and his colleagues estimated the cost by adding up grants that funded papers that were

retracted for misconduct, and grants to authors censured for misconduct by the Office of Research Integrity. The total is less than 1% of the NIH budget over the 20 years.

## Reef under threat

The outlook is poor for the iconic Great Barrier Reef off the coast of Queensland, Australia, according to a report released last week. Climate change is the greatest threat, as rising sea temperatures increase the risk of coral bleaching and water becomes more acidic. Nutrient and pesticide pollution from land, overfishing and destruction of coastal habitat continue to be problems, says the Great Barrier Reef Marine Park Authority. In April, the reef narrowly missed being listed as ‘in danger’ when the United Nations Educational, Scientific and Cultural Organization (UNESCO) considered the conservation status of its world heritage sites.

## Tobacco control

European legislation to regulate tobacco was shaped by the tobacco industry, according to a study published on 13 August (H. Costa *et al.* *Tob. Control* <http://doi.org/t75>; 2014). Researchers scrutinized draft texts of proposed revisions to Europe's tobacco-regulation law and found that changes to texts between

2010 and 2013 resembled statements from the tobacco industry much more closely than those of health groups. Aggressive lobbying from the industry was “associated with significant policy shifts” towards the tobacco industry, such as the scaling back of proposed limits on where cigarettes could be sold, the paper adds. The revised legislation was adopted in March 2014.

## FUNDING

## BRAIN grants

On 18 August, the US National Science Foundation awarded 36 small grants totalling US\$10.8 million to projects in President Barack Obama's Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative. The agency originally planned to fund about 12 grants, but decided to triple that number after receiving nearly 600 applications. Selected projects include models that help computers to recognize different parts of and patterns in the brain. Nearly all projects were awarded the maximum amount of \$300,000 over two years. See [go.nature.com/qwhwld](http://go.nature.com/qwhwld) for more.

## BUSINESS

## Insomnia drug

The US Food and Drug Administration approved on 13 August the first drug for insomnia that works by interfering with the signalling of neurotransmitters called orexins, which regulate wakefulness. The drug, called suvorexant (Belsomra), is made by Merck of Whitehouse Station in New Jersey, and will be available in four strengths. The company had originally petitioned for approval of higher doses, but was rebuffed last year by regulators amid concerns about the risk of next-day drowsiness.

**NATURE.COM**

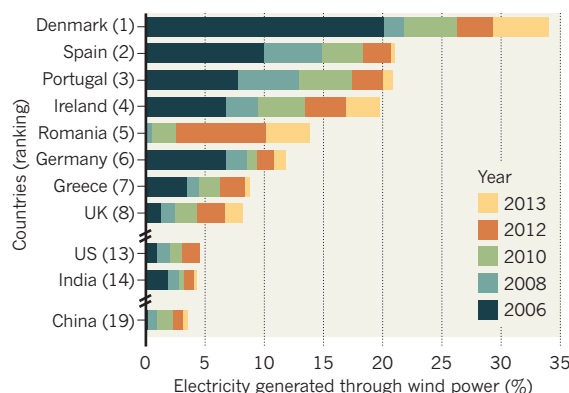
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

China and the United States have the greatest installed wind-power capacity of any country, according to an 18 August report from the US Department of Energy. But the United States added just 1.1 gigawatts last year, placing it sixth in annual growth. The United States ranks even lower in terms of how much of the energy that it consumes comes from wind power (see graph). US firms had little motivation to finish plants in 2013 because projects only had to start construction by the year's end to qualify for tax credits.

## BETTING ON WIND POWER

Denmark has been generating a greater proportion of its electricity through wind power than most other countries for several years.



# NEWS IN FOCUS

**CLIMATE CHANGE** Report reveals warming threats to Tibetan Plateau **p.240**

**ARCHAEOLOGY** Neanderthals died out in Europe much earlier than thought **p.242**

**POLICY** NIH to probe racial bias in grant awards **p.243**



**ANTARCTIC SCIENCE** Lake beneath the ice yields microbial garden **p.244**

GYGAIA PROJECTS



US archaeologist Chris Roosevelt can use a drone for research in Turkey, but home is a no-fly zone.

## POLICY

# US drone research hits regulatory turbulence

*Federal rules ground scientists using remotely piloted aircraft at private universities.*

BY HEIDI LEDFORD

Over the past few months, archaeologist Chris Roosevelt has been remotely piloting a petite six-rotor helicopter named Hydra over western Turkey, snapping high-resolution photographs of the scrubby terrain. The pictures will be assembled into three-dimensional maps to help Roosevelt and his colleagues to understand how people who built a kingdom in the region 3,500 years ago interacted with the landscape. But back home

at Boston University in Massachusetts, he is not allowed to fly the copter — not even to teach budding archaeologists on campus how to use it.

On 23 June, the US Federal Aviation Administration (FAA) told drone users that such aircraft cannot be flown for ‘commercial’ use — defined as any activity that is not recreational or part of a government programme. That includes research and teaching at private, but not public, universities. Even though the announcement merely clarified existing rules, it caught many by surprise and reinforced a seemingly arbitrary

dichotomy that treats researchers differently on the basis of their institutional affiliation.

“It would be the rare person to think of archaeological activities as a commercial enterprise,” says Roosevelt. “But for the FAA, commercial automatically includes education and research.”

Drones’ small size, low cost and long-lasting batteries have made them popular in research and business. Scientists deploy the machines to study everything from atmospheric science to zoology, and companies are eager to harness ▶



► them for aerial photography and package delivery. This has put the FAA under pressure to develop regulations prohibiting uses of drones that compromise safety or violate privacy. A few years ago, the US Congress ordered the FAA to develop such rules by September 2015. But in a report this July, the US Department of Transportation's inspector general expressed doubt that the FAA would make its deadline.

The administration has shut down almost all commercial drone flights until those regulations are in place — to the ire of some researchers. “The FAA's severe restrictions on research and education are out of all proportion to the actual risks posed by small-model aircraft used responsibly on private property,” says Paul Voss, an engineer who does atmospheric research at Smith College in Northampton, Massachusetts. On 25 July, Voss and 28 of his colleagues wrote to the FAA outlining their concerns. Because the agency is forbidden by law to put such severe restrictions on model aircraft flown by hobbyists, they say, it has established a bizarre double standard in which recreational users can pilot drones but trained researchers cannot.

Geographer Scott Drzyzga of Shippensburg University in Pennsylvania, for example, enjoys flying small helicopters in his free time, and had hoped to show his students how to use them to capture geographical information. “There are things I can do as a recreational user that I cannot do in front of my students,” he says. “It's frustrating.”

Alison Duquette, a spokesperson for the FAA, says that the distinction has been in place for years, and the agency has always made it clear.

Engineer Mary Cummings at Duke University in Durham, North Carolina, hopes to overcome the regulatory obstacle by teaming with colleagues at a nearby public university to test drones for wildlife tracking. But for now the FAA rules confine her to indoor testing — poor conditions for determining whether her machines can follow animals in the wild. “We are looking to go to another country to address this problem,” she says.

Even at public universities, red tape and legal fears have spooked researchers who want to use drones. Larry Purcell, a crop scientist at the University of Arkansas in Fayetteville,

has used the devices to monitor the effects of drought on soya beans. He did not know that he was required to fill out a 50-page application for each field he wanted to survey. Instead, he sought and obtained verbal permission from FAA staff at a local airport. Shortly after his university issued a press release about his work in 2011, Purcell received a cease-and-desist letter from the FAA's regional office. “It's still very confusing as to what's allowed right now and what's not,” he says. Purcell has turned to kites to carry his cameras aloft, but the compromise has its shortcomings: no wind means no data.

Others at both public and private institutions have continued their work without notifying the agency. Several scientists contacted by *Nature* declined to describe their work with drones, for fear of drawing FAA attention.

And some have gone to extremes to stay out of trouble. A few years ago, Voss designed 200-gram drones small enough to pack into a suitcase for field work in Antarctica. The craft worked well, but he has since got rid of them. “I don't even want them in the lab,” he says. “I want to be squeaky clean.” ■ [SEE EDITORIAL P.231](#)

## ENVIRONMENT

# Double threat for Tibet

*Climate change and human development are jeopardizing the plateau's fragile environment.*

BY JANE QIU IN LHASA

A comprehensive environmental assessment of the Plateau of Tibet has found that the region is getting hotter, wetter and more polluted, threatening its fragile ecosystems and those who rely on them.

The plateau and its surrounding mountains cover 5 million square kilometres and hold the largest stock of ice outside the Arctic and Antarctic; the region is thus often referred to as the Third Pole. And like the actual poles, it is increasingly feeling the effects of climate change, but rapid development is putting it doubly at risk, the report says.

Released in Lhasa on 9 August by the Chinese Academy of Sciences (CAS) and the government of Tibet, the assessment aimed to address gaps in knowledge about the extent of the problems the 4,500-metre-high plateau faces. It finds that precipitation has risen by 12% since 1960, and temperatures have soared by 0.4°C per decade — twice the global average.

In addition, glaciers are shrinking rapidly and one-tenth of the permafrost has thawed in the past decade alone. This means that the number of lakes has grown by 14% since 1970, and more than 80% of them have expanded since, devastating surrounding pastures and communities.



The plateau feeds Asia's biggest rivers (see 'Running wild'), so these problems are likely to affect billions of people, the report says. Pollution from human and industrial waste as a result of rapid development is also a serious risk.

But the assessment also suggests ways to combat the problems, calling on the Chinese and Tibetan governments to make conservation and environmental protection top priorities. It will help in the design of “policies for mitigating



KIERAN DODDS/PANOS



Hot, dry weather and progressive urbanization are turning grasslands into sand near the headwaters of the Yellow, Yangtze and Mekong rivers.

climate change and striking a balance between development and conservation”, says Meng Deli, Tibet’s vice-chairman.

“The Tibetan plateau is getting warmer and wetter,” says Yao Tandong, director of the CAS Institute of Tibetan Plateau Research in Beijing, who led the assessment. This means that vegetation is expanding to higher elevations and farther north, and growing seasons are getting longer. But some areas, such as the headwater region of Asia’s biggest rivers, have become warmer and drier and are being severely affected by desertification and grassland and wetland degradation.

Human activity, too, is on the rise. The population of the plateau reached 8.8 million in 2012, about three times higher than in 1951. And the number of livestock has more than doubled, putting more strain on grasslands.

### MULTIPLE MENACES

Growing urbanization is creating more waste than the region can handle. Tibet has the capacity to treat 256,000 tonnes of domestic solid waste a year, less than the amount generated by its two largest cities, Lhasa and Shigatse. “You see a lot of rubbish lying around the plateau, including headwater regions,” says Kang Shichang, a glaciologist at the CAS Institute of Cold and Arid Regions Environmental and

Engineering Research Institute in Lanzhou. “It’s an environmental menace.”

A bigger threat comes from mining. According to the assessment, Tibetan mines produced 100 million tonnes of wastewater in 2007 and 18.8 million tonnes of solid waste in 2009. Because most of the mines are open pits and have limited environmental oversight, “air, water and soil pollution is particularly serious”, says the report. Officials release few details about actual pollution levels.

Pollution is coming not just from local sources. Dust, black carbon, heavy metals and other toxic compounds are being blown in from Africa, Europe and southern Asia. The dust and carbon residues are darkening glaciers, making them more susceptible to melting, and the toxic chemicals are poisoning crops, livestock and wildlife.

But the threats from mining and pollution are dwarfed by the potential repercussions of changes in ice and vegetation cover, the assessment says. Different surfaces — snow, grassland, desert — reflect and absorb different amounts of solar radiation, affecting how the air above them is heated. This means that changes in coverage are likely to affect the onset and strength

of Asian monsoons. It also has important ramifications for the livelihood of downstream river communities because the glaciers, permafrost and ecosystems act as a giant sponge, helping to control the release of water and prevent floods. “The significance of the assessment goes beyond national borders,” says David Molden, head of the International Centre for Integrated Mountain Development in Kathmandu.

Temperatures in the plateau are projected to rise by between 1.7°C and 4.6°C by the end of 2100 compared with the 1996–2005 average, based on the best- and worst-case global-emissions scenarios. So as urbanization and climate change tighten their grip, researchers worry that unbridled development will devastate the plateau’s environment. To protect it, the report says, the central government must evaluate local officials on the basis of their environmental, not just economic, achievements. It must also invest more in ecological compensation, for example by paying herders more to cut their livestock numbers. Moreover, it must be much more open about pollution incidents.

“Tibet will be a test case of how seriously China takes ecological protection,” says Yao. “Safeguarding the plateau environment is crucial not only for sustainable development of the region, but also to social stability and international relations.” ■

**“Tibet will be a test case of how seriously China takes ecological protection.”**

### MORE ONLINE

#### Q & A

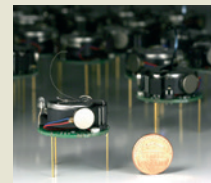


NIH director Francis Collins on the changing face of biomedical research [go.nature.com/53vzmj](http://go.nature.com/53vzmj)

#### MORE NEWS

- Dust particles from outside the Solar System found in probe’s haul [go.nature.com/wkvtnv](http://go.nature.com/wkvtnv)
- Embalming predates pharaohs by 1,000 years [go.nature.com/3toihr](http://go.nature.com/3toihr)
- Eight-year sentence sought in thesis copyright case [go.nature.com/rxhuik](http://go.nature.com/rxhuik)

#### VIDEO OF THE WEEK



Watch 1,000 robots self-organize into a swarm [go.nature.com/mtznhq](http://go.nature.com/mtznhq)

ANDREW HARRER/  
BLOOMBERG/GETTYMICHAEL RUBENSTEIN,  
HARVARD UNIV.



Archaeologist Tom Higham has dated Neanderthal jawbones and other samples from across Europe.

## ARCHAEOLOGY

# Bone technique redrafts prehistory

*Carbon-dating improvements show that Neanderthals disappeared from Europe much earlier than thought.*

BY EWEN CALLAWAY

Neanderthals and humans lived together in Europe for thousands of years, concludes a timeline based on radiocarbon dates from 40 key sites across Europe. The results<sup>1</sup>, published today in *Nature*, may help to end a century-old deadlock over the demise of the Neanderthals and their relationship to humans.

The researchers used 196 radiocarbon dates of organic remains to show that Neanderthals disappeared from Europe around 40,000 years ago, but still long after humans arrived in the continent. "Humans and Neanderthals were living contemporaneously for quite some period of time in different parts of Europe," says Tom Higham, an archaeologist at the University of Oxford, UK, who led the study. The long overlap provided plenty of time for cultural exchange and interbreeding, he adds.

Exactly what happened 30,000–50,000 years ago still vexes archaeologists because the period is right at the limit of accurate radiocarbon dating. The technique is based on measuring the steady loss of radioactive carbon-14 molecules in organic remains. But after 30,000 years, 98% of the isotope

is gone and younger carbon molecules are starting to infiltrate bones, making remains seem younger than they are. This means that dates for the final Neanderthals and for the first human occupations of Europe have been unreliable, fomenting the debate.

But over the past decade, Higham and his team have developed techniques that provide more accurate readings in bones up to 55,000 years old (see *Nature* **485**, 27–29; 2012). First, they use a chemical pretreatment to remove the contaminating carbon from the collagen in bones, then they measure the minuscule amounts of radiocarbon using a particle accelerator.

The technique has allowed the researchers to redraft the prehistory of Europe cave by cave, and show that early humans arrived in southwestern England<sup>2</sup> and Italy's 'heel'<sup>3</sup>, for example, well over 40,000 years ago. They have now applied the technique to Neanderthal occupations across Europe, which are associated with stone tools known as Mousterian artefacts. From the Black Sea to the Atlantic coast of France, these artefacts

and Neanderthal remains disappear from European sites at roughly the same time, 39,000–41,000 years ago, Higham's team conclude. The data challenge arguments that Neanderthals endured in refuges in the southern Iberian Peninsula until as recently as 28,000 years ago<sup>4</sup>.

Humans, argues Higham's team, were in Italy as early as 45,000 years ago, where they developed a stone-tool culture known as the Uluzzian industry. The team estimates that humans and Neanderthals overlapped for up to 5,400 years in parts of southern Europe, yet to a much lesser extent or not at all in other parts of the continent. "They were certainly in the same areas," Higham says.

The coexistence also supports a controversial idea that some Neanderthal artefacts, such as shell beads and stone tools of the Châtelpéronian industry that appeared in France and Spain more than 40,000 years ago, emerged from contact with humans, Higham says.

## PLENTY OF TIME

"I absolutely agree with Tom," says Paul Mellars, an archaeologist at the University of Cambridge, UK, who has long championed the idea that Neanderthals borrowed technologies from humans. "There would be opportunities for contacts or interactions thousands of times in most, if not all, areas of Europe," he says.

Others are more sceptical. Clive Finlayson, director of the heritage division at the Gibraltar Museum whose team dated the 28,000-year-old Neanderthal charcoal remains from the tip of Gibraltar<sup>4</sup>, questions the timeline's sweeping conclusions. Archaeologists are unlikely to ever find the last Neanderthal occupation, he argues, and the methods that Higham's team used to remove contamination do not work well for bones from warmer sites because collagen is not preserved as well there as in cooler sites. This would be the case in southern Iberia, where Finlayson believes the last Neanderthals lived. "I'm hugely worried that we're building a castle in the air here," he says.

Meanwhile, Higham hopes that the timeline will address other mysteries surrounding Neanderthals, such as why they died out and how they interacted with humans. DNA recovered from remains in Europe and western Asia, for example, show that humans and Neanderthals interbred more than 50,000 years ago, probably as the common ancestor of Europeans and Asians emerged from Africa.

There is still no evidence that humans and Neanderthals interbred while in Europe, but thousands of years of overlap makes sex more likely, Higham says. "I do like the idea that they aren't really extinct and they do live on in us." ■

► **NATURE.COM**  
Watch an animation  
of the Neanderthal  
timeline:  
[go.nature.com/b4zt8c](http://go.nature.com/b4zt8c)

1. Higham, T. *et al. Nature* **512**, 306–309 (2014).
2. Higham, T. *et al. Nature* **479**, 521–524 (2011).
3. Benazzi, S. *et al. Nature* **479**, 525–528 (2011).
4. Finlayson, C. *et al. Nature* **443**, 850–853 (2006).



## POLICY

# NIH to probe racial disparity in grant awards

US agency will assess whether grant reviewers are biased against minority applicants.

BY SARA REARDON

**R**ichard Nakamura, director of the Center for Scientific Review at the US National Institutes of Health (NIH), does not consider himself to be racially biased. Yet a test of his speed at associating certain words with faces of different races revealed a slight unconscious prejudice against minorities. If the director of the institute that oversees the NIH's grant process harbours these inclinations, he wonders, are grant reviewers affected as well?

To answer that question, the NIH will launch ambitious analyses beginning in September to determine whether bias hampers minority scientists who seek agency funding. A 2011 study in *Science* found that white researchers receive NIH grants at nearly twice the rate that African American researchers do (see 'Grant gap'). Even when factors such as publication record and training are considered, an African American scientist is still only two-thirds as likely as a white scientist to be funded (D. K. Ginther *et al. Science* **333**, 1015–1019; 2011). The disparity seems to arise early during the review process, when grants are first rated.

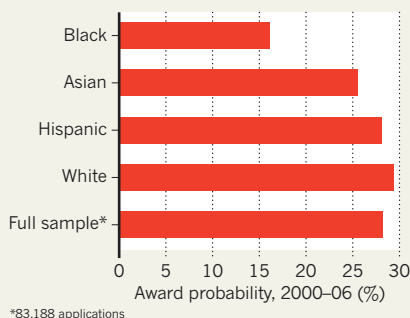
The findings spurred the NIH to launch a ten-year, US\$500-million effort in 2012 to train and mentor minority scientists. But officials acknowledge that the racial gap among grantees is not just because there are fewer qualified applications from minority researchers. Now the agency will look inward to determine where its grant process may be failing — and what to do about it.

One basic issue that the NIH will address is whether grant reviewers are thinking about an applicant's race at all, even unconsciously. A team will strip names, racial identification and other identifying information from some proposals before reviewers see them, and look at what happens to grant scores. (Such identity stripping is surprisingly difficult: even citations might reveal who the applicant is, and reviewers need some information about an applicant to make a fair appraisal.) The results could be telling. "If the disparity drops with anonymization, that's clear evidence of bias," says Nakamura.

Such a finding would be in line with other results in this area. A study published this year found that faculty members in US universities are less likely to respond to interview requests from prospective students whose names are

## GRANT GAP

Scientists from racial minority groups are less likely than white applicants to receive research funding from the US National Institutes of Health.



associated with minority groups than they are to identical requests from students with 'white' names (K. L. Milkman *et al. Soc. Sci. Res. Network* <http://doi.org/t9h>; 2014).

The NIH will also study reviewers' work in finer detail, by analysing successful applications for R01 grants, the NIH's largest funding programme for individual investigators. The goal is to see whether researchers can spot trends in the language used by reviewers to describe proposals put forward by applicants of different races. There is precedent for detectable differences: in a paper to be published in *Academic Medicine*, a team led by Molly Carnes, a physician at the University of Wisconsin-Madison, used automated text analysis to show that reviewers' critiques of R01 grant applications by women tended to include more words denoting praise, as though the writer is surprised at the quality of the work. And numerous other studies show that different standards exist for men and women in a variety of fields. "Women do, indeed, have to be twice as good to get the same competence rating as a man," says Carnes.

The NIH will also analyse text in samples of reviewers' unedited critiques. The Center for Scientific Review typically edits the wording and grammar of these reviews before grant proposals are returned to applicants, but even the subtlest details of such raw comments might hold clues about bias. Nakamura says that reviewers will not be told whether their comments will be analysed, because that in itself would bias the sample. "We want them to be sloppy," he says.

The NIH's Study Sections, in which

review groups discuss the top 50% of grant applications, might also harbour bias: the 2011 *Science* paper found that submissions authored by African Americans are less likely to be discussed in the meetings. But when they are, a negative comment arising from even one person's unconscious bias could have a major impact in such a group setting, says John Dovidio, a psychologist at Yale University in New Haven, Connecticut, and a member of the NIH's Diversity Working Group. "That one person can poison the environment," he says.

Even if the NIH investigation does not turn up evidence of bias, it may still reveal some of the causes of the racial disparity in the NIH's grant-making process. Perhaps grants from minority researchers are more likely to be written in a way that does not appeal to reviewers, says Monica Basco, executive secretary of the Diversity Working Group's peer-review subcommittee. That would suggest fixes such as grant-writing help. Evidence of bias would be harder to address, and any interventions would need to be tailored to address the point at which it occurs, says Basco.

Nakamura expects that the NIH's effort to identify and root out prejudice, which he says could cost up to \$5 million over three years, might prove controversial. "People resent the implication they might be biased," he says — an idea borne out by some responses to his 29 May blogpost on the initiative. One commenter wrote, "It is absolutely insulting to be accused of review bigotry. Please tell me why I should continue to give up my time to perform peer review?"

But Nakamura believes that the NIH — and reviewers — need to keep open minds. After all, he says, "we are human beings with emotions and feelings we're not in control of". ■ [SEE EDITORIAL P.231](#)

## CORRECTIONS

The News Feature 'Scientists and the social network' (*Nature* **512**, 126–129; 2014) gave the wrong affiliation for Laura Warman, who is at the University of Hawaii at Hilo. And the News story 'Health check for deep-sea mining' (*Nature* **512**, 122–123; 2014) gave an incorrect list of the resources found around deep-sea vents. In fact, only polymetallic sulphides are common near the vents.

SOURCE: D. K. GINTHER *ET AL. SCIENCE* **333**, 1015–1019 (2011)




# ANTARCTICA'S SECRET GARDEN

**SAMPLES FROM A LAKE  
HIDDEN UNDER 800 METRES  
OF ICE CONTAIN THOUSANDS  
OF MICROBES AND HINT AT  
VAST ECOSYSTEMS YET TO  
BE DISCOVERED.**

**BY DOUGLAS FOX**







A cold breeze blew off the Antarctic plain, numbing the noses and ears of scientists standing around a dark hole in the ice. Flecks of ice crackled off a winch as it reeled the last few metres of cable out of the hole. Two workers in sterile suits leaned over to grab the payload — a cylinder the length of a baseball bat — dangling at the end of the cable. They used a hammer to chip away the ice and a blow drier to thaw part of the assembly. “Did it close?” asked the winch operator.

“Yeah,” shouted John Priscu, a microbial ecologist from Montana State University in Bozeman. The cylinder rested heavily in his gloved hands — evidence that it had filled with water and sealed shut before its long journey to the surface. The fluid inside came from one of the most isolated bodies of water on Earth: Lake Whillans, trapped beneath 800 metres of ice just 640 kilometres from the South Pole. Hardly a word was spoken as Priscu hefted the vessel against his shoulder and shuffled into a metal shipping container, where the team had set up a cramped, makeshift laboratory.

That water, obtained on 28 January 2013, was the first sample ever retrieved directly from a subglacial lake. Although Priscu and other scientists had long yearned to explore Antarctica’s hidden lakes and look for resident life, efforts to drill into them have been stymied by the threat of contamination, which would cast doubt on any life found and could introduce invasive organisms into the lake. Priscu and his team spent six years devising safe sampling procedures, and then had to surmount numerous logistical hurdles, such as transporting hundreds of tonnes of equipment to the remote site.

The researchers have been studying the samples since they reached the lake and have found that an abundance of life lurks beneath Antarctica’s blanket of ice. In this week’s issue of *Nature*<sup>1</sup>, Priscu and his team report finding 130,000 cells in each millilitre of lake water — a density of microbial life similar to that in much of the world’s deep oceans<sup>2</sup>. And with nearly 4,000 species of bacteria and archaea, the community in the lake is much more complex than might be expected from a world sealed off from the rest of the planet. “I was surprised by how rich the ecosystem was,” says Priscu. “It’s pretty amazing.”

Samples from the lake show that life has survived there without energy from the Sun for the past 120,000 years, and possibly for as long as 1 million years. And they offer the first look at what may be the largest unexplored ecosystem on Earth — making up 9% of the world’s land area. “There’s a thriving ecosystem down there,” says David Pearce, a microbiologist at Northumbria University, UK, who was part of a team that tried, unsuccessfully, to drill into a different subglacial body, Lake Ellsworth, in 2013. “It’s the first time that we’ve got a real insight into what organisms might live beneath the Antarctic continent,” he says.

#### LIFE ON ICE

The ice above Lake Whillans is mind-numbingly flat, making it nearly impossible to imagine that anything unusual hides beneath it. I first travelled there in 2007 as a journalist covering a scientific expedition to the lake, which had been discovered earlier that year through remote satellite measurements. I returned in January 2013, embedded as a reporter with the team that Priscu

led with two other scientists to sample the lake. That project, called Whillans Ice Stream Subglacial Access Research Drilling, involved collaboration between nearly two dozen researchers from 15 universities across five countries. The US National Science Foundation had invested roughly US\$20 million in the effort, which included building a hot-water drill to get into the lake without contaminating it.

The idea that lakes might lurk beneath Antarctica’s frozen cover was not widely considered until the 1990s, when ice-penetrating radar and seismic mapping yielded the first solid evidence of subglacial lakes. Nearly 400 are now known. They are fed by water that melts from the base of the ice sheet at rates of a few millimetres per year, caused by ambient heat from deep within the planet (see ‘Invisible lakes’).

**“It’s the first time  
that we’ve got a  
real insight into  
what organisms  
might live beneath  
the Antarctic  
continent.”**

Lake Whillans resembles nothing on Earth’s surface. The weight of the ice forces the subglacial water upwards, causing the lake to sit at a slant on the side of a hill. It is a thin lens of water — only 2 metres deep and nearly 60 square kilometres in area — held in a pocket of low pressure created by the thinning of the ice sheet as it oozes over the hill.

The drill camp materialized on this lonely frontier in January 2013, when tractors arrived pulling shipping containers on massive skis. In their two-week journey from the coast, the tractors hauled 500,000 kilograms of gear and fuel, mobile labs, a machine shop and a hot-water drill that filled six cargo containers. Within two weeks, the camp was a noisy, industrial place, populated by three dozen people, a flock of tents flapping in the steady breeze and two roaring 225,000-watt generators. The polar summer resembled a mild winter in Minneapolis, Minnesota, with temperatures 5–15°C below freezing.

It took seven days to drill through the ice sheet. To prevent contamination of the lake, the crew used ultraviolet radiation, water filtration and hydrogen peroxide to sterilize the machinery and the water used to bore through the ice. As the team neared the lake, progress slowed to a crawl when difficulty in steering the drill bedevilled the crew for an agonizing 36 hours.

At 7:30 a.m. on 27 January, a voice crackling through a handheld radio summoned me to the drill control room. Inside, six ice drillers in overalls stared at a computer screen showing a line shooting upwards on a graph, indicating that the water in the borehole had risen 28 metres, pushed up by a gush of water from the lake below. The lake was a balmy –0.5°C, warmer than the drill camp that day.

The researchers pulled up the first sample the next day. Within minutes of raising the grey vessel, they decanted its contents: a honey-coloured broth that turned out to be richer in minerals than anyone had expected. The first cells were spotted several hours later under a microscope — green dots lit up by DNA-sensitive dye. Tests done over the next several days confirmed that those cells were alive. Twenty scientists and graduate students worked around the clock to collect 30 litres of water and several sediment cores from the lake. Before the hole froze shut, the team also measured the water chemistry in the lake and geothermal heat flowing up through the sediments. Sample boxes accumulated in a cave dug out of the snow on the edge of camp.

Over the past year, researchers have worked with those samples to assemble a portrait of life beneath the ice sheet. They have isolated and grown cultures of about a dozen species of microbe. And DNA sequencing has revealed

John Priscu  
carries a sample  
from subglacial  
Lake Whillans.

J. T. THOMAS

signs of 3,931 species in all — many of them related to known microbes that break down minerals for energy.

Although contamination is always a concern, researchers not connected with the Lake Whillans project say that the sterilization precautions seem to have worked well. One sign is that the microbial density of the drilling water in the hole was 200 times lower than that of the lake samples, says Peter Doran, an Earth scientist at the University of Illinois in Chicago, who worked with the US National Research Council for ten years to develop guidelines for sampling Antarctic lakes cleanly. Doran was convinced by the evidence of diverse microbial life in the lake. “They found it in such a way that it can’t be questioned. It’s pretty iron-clad,” he says.

### VITAL SIGNS

Overall, life in Lake Whillans works much like ecosystems at the surface, but its deep denizens do not have access to sunlight and so cannot rely on photosynthesis for the energy needed to fix carbon dioxide dissolved in the lake water.

The genetic analyses by the team show that some of the lake’s microbes are related to marine species that derive energy by oxidizing iron and sulphur compounds from minerals in sediment. But according to the DNA data, the lake’s most abundant microbes oxidize ammonium, which is likely to have a biological origin.

“The ammonium is probably a relic of old marine sediments,” says Priscu, referring to dead organic matter that accumulated millions of years ago when the region was covered by shallow seas rather than glaciers.

Only single-celled bacteria and archaea have turned up in samples from Lake Whillans — but the particular DNA tests used so far were not designed to detect other types of organism. This preserves the possibility that Lake Whillans might yet be found to harbour more complex life, such as protozoa — or even sub-millimetre animals such as rotifers, worms or eight-legged tardigrades, all known to live in other parts of Antarctica. Air bubbles in the overlying ice supply oxygen to the lake, so that is not a limiting factor. But the low rate of carbon fixation by microbes might provide too little food for multicellular life.

Lake Whillans receives about one-tenth the amount of new carbon per square metre per year as the world’s most nutrient-starved ocean floors, which support sparse animal populations. Although the chances are slim that Priscu and his colleagues will find animals in Lake Whillans, they plan to look for them using better-tailored DNA assays. For now, the researchers are puzzling over the origins of the microbial residents of the lake.

➔ **NATURE.COM**

For a video of the Lake Whillans expedition, see:

[go.nature.com/pb3jok](http://go.nature.com/pb3jok)

## INVISIBLE LAKES

Researchers have tried to drill into three lakes under Antarctica’s ice sheet. To collect samples from Lake Whillans, a team used tractors to bring equipment from a US base on the coast.



The big question is whether Antarctica’s subglacial communities are made of ‘survivors’ or ‘arrivers’.

Survivors would be the descendants of microbes that lived in the sediments when the area was covered by open ocean, as it has been periodically over the past 20 million years. Alternatively, Lake Whillans might be populated by wind-blown microbes — the ‘arrivers’ — that were deposited on the ice and worked their way down over 50,000 years as ice melted off the bottom of the glacier.

It is also possible that some organisms entered the lake more recently, carried in by sea water seeping under the ice sheet. Lake Whillans sits just 100 kilometres from the grounding line, where the ice sheet transitions from resting on ground to floating on the ocean. That line shifts as the ice thins and thickens, so it is possible that the lake exchanged water — and microbes — with the ocean during the past few thousand years, says Christina Hulbe, a glaciologist at the University of Otago in Dunedin, New Zealand, who has long studied that area of Antarctica.

Other findings from the lake samples have led to some tantalizing ideas. Traces of fluoride in its water offer possible evidence of hydrothermal vents in the area — rich sources of chemical energy that have the potential to support islands of exotic life, such as worms or heat-loving microbes. “It’s probable that there are hydrothermal systems in there,” says Donald Blankenship, a glaciologist at the University of Texas at Austin. The lake occupies a broad rift valley where Earth’s crust has thinned, and radar surveys by Blankenship show putative volcanoes under the ice<sup>3,4</sup>.

The results emerging from Lake Whillans could also shed light on how Antarctica influences the nearby ocean and even the entire world. If microbes beneath the ice sheet play an important part in altering the minerals in the sediments, as the latest data suggest, those organisms might supply iron to the subglacial

waters that eventually reach the ocean.

This process could provide an important source of nutrients to the chronically iron-starved ecosystems in the Southern Ocean<sup>5</sup>, says Martyn Tranter, a marine biogeochemist the University of Bristol, UK.

In addition, the presence of small amounts of a chemical called formate in the water of Lake Whillans suggests the possibility that methane, a potent greenhouse gas, is produced in the deeper, oxygen-poor sediments beneath the lake. A 2012 study estimated that the sediments under the Antarctic ice sheet contain hundreds of billions of tonnes of methane — a reservoir equal to that stored in the Arctic’s permafrost — which could potentially escape and exacerbate global warming if the ice retreats<sup>6</sup>.

Lake Whillans provides only a local snapshot of life beneath the ice, and several teams are trying to fill in the picture by exploring other subglacial lakes. A Russian team is now analysing water from Lake Vostok, a lake in a deep tectonic rift in eastern Antarctica that is covered by 3.7 kilometres of ice. Researchers say that analysing those samples presents challenges because the water spent a year frozen in the bottom of the borehole before being brought to the surface. And as the ice was raised, it was exposed to the kerosene drilling fluid in the borehole.

Closer to Lake Whillans, Pearce and his colleagues attempted in 2013 to drill into Lake Ellsworth, which sits under 3.4 kilometres of ice in a glacial fjord, but they were forced to abandon the effort after difficulties arose with steering the drill.

With its thinner ice covering, Lake Whillans was an easier target than Ellsworth or Vostok, but it did not give up its secrets easily. The day after the first sample was retrieved, a camera lowered into the hole presented a mesmerizing scene as it neared the lake. Iridescent flakes of ice drifted upwards — a snow shower in reverse, and a sign that the hole was quickly refreezing. The scientists’ instruments soon began to catch in the narrowing hole, forcing the drillers to pump in hot water to widen it. This tug-of-war lasted for four days before the team abandoned the hole to its inevitable fate, broke camp and flew their hard-won samples home. ■

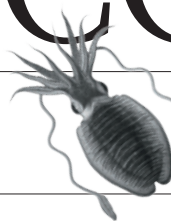
**Douglas Fox** is a freelance journalist in Northern California.

1. Christner, B. C. *et al. Nature* **512**, 310–313 (2014).
2. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
3. Blankenship, D. D. *et al. Nature* **361**, 526–529 (1993).
4. Schroeder, D. M., Blankenship, D. D., Young, D. A. & Quartini, E. *Proc. Natl Acad. Sci. USA* **111**, 9070–9072 (2014).
5. Death, R. *et al. Biogeosciences* **11**, 2635–2643 (2014).
6. Wadhwa, J. L. *et al. Nature* **488**, 633–637 (2012).



# COMMENT

**HISTORY** Reappraising Aristotle as the first scientist **p.250**

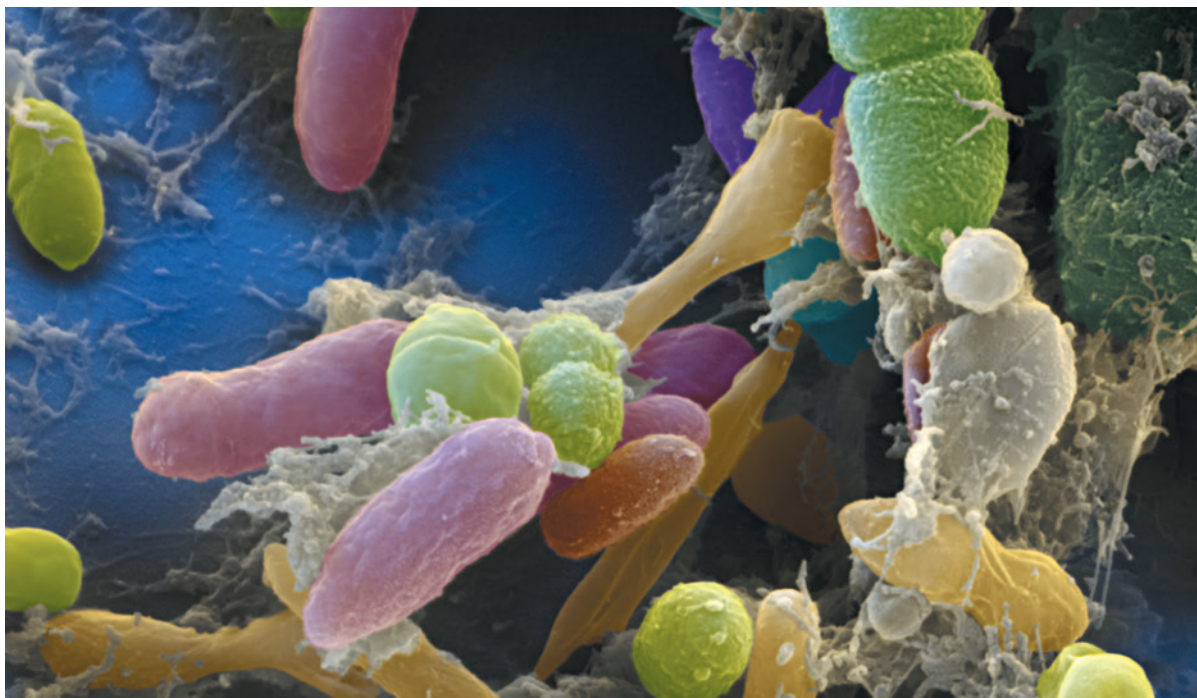


**SPACE** Exploring our nearest neighbour, Venus **p.252**

**POLICY** Proposed definition of 'invasive' threatens native biodiversity **p.253**

**AGEING** A call for more social science on the cultural aspects of longevity **p.253**

EYE OF SCIENCE/SPL



A scanning electron micrograph of bacteria in human faeces, in which 50% of species originate from the gut.

## Microbiome science needs a healthy dose of scepticism

To guard against hype, those interpreting research on the body's microscopic communities should ask five questions, says **William P. Hanage**.

Explorations of how the microscopic communities that inhabit the human body might contribute to health or disease have moved from obscure to ubiquitous. Over the past five years, studies have linked our microbial settlers to conditions as diverse as autism, cancer and diabetes.

This excitement has infected the public imagination. 'We Are Our Bacteria', proclaimed one headline in *The New York Times*. Some scientists have asserted that antibiotics are causing a great 'extinction' of the microbiome, with dire consequences for human health<sup>1</sup>. Companies offer personalized

analysis of the microbial content of faecal samples, promising consumers enlightening information. Separate analyses from the same person can, however, vary considerably, even from the same stool sample. Faecal transplants have been proposed — some more sensible than others — for conditions ranging from diabetes to Alzheimer's disease. With how-to instructions proliferating online, desperate patients must be warned not to attempt these risky procedures on themselves.

Microbiomics risks being drowned in a tsunami of its own hype. Jonathan Eisen, a microbiologist and blogger at the University

of California, Davis, bestows awards for "overselling the microbiome"; he finds no shortage of worthy candidates.

Previous 'omics' fields have faltered after murky work slowed progress<sup>2</sup>. Technological advances that allowed researchers to catalogue proteins, metabolites, genetic variants and gene activity led to a spate of associations between molecular states and health conditions. But painstaking further work dampened early excitement. Most initial connections were found to be spurious or, at best, more complicated than originally believed.

The history of science is replete with ►

► examples of exciting new fields that promised a gold rush of medicines and health insights but required scepticism and years of sloggling to deliver even partially. As such, the criteria for robust microbiome science are instructive for all researchers. As excitement over the microbiome has filtered beyond academic circles, the potential mischief wrought by misunderstanding encompasses journalists, funding bodies and the public.

### CRUCIAL QUESTIONS

Here are five questions that anyone conducting or evaluating this research should ask to keep from getting carried away by hype.

**Can experiments detect differences that matter?** Profiling a microbiome could produce a catalogue at the level of phyla, species or genes. Much work relies on analysis of *16S rRNA*, an ancient gene that tolerates little variation and so is reliably found across the bacterial kingdom. But this allows only a coarse sorting. For example, microbiomes associated with obesity have been distinguished by different ratios of bacterial phyla, which encompass a staggering range of diversity. If this criterion were used to characterize animal communities, an aviary of 100 birds and 25 snails would be considered identical to an aquarium with 8 fish and 2 squid, because each has four times as many vertebrates as molluscs. Even within a single species, strains often differ greatly in the genes they contain.

Modern technology now allows for finer distinctions: we can study more genes in a sample, an ability that may enable us to decipher 'metabolic networks' revealing the biochemical reactions that a microbiome can perform. This kind of analysis could identify gene combinations, potentially from multiple species across a microbial community, that affect health for good or ill. However, pinning an outcome to any particular entity is likely to be hard unless the networks are already well characterized.

To take a simple example from a single bacterial species, we could show that vaccination eliminated 30% of known pneumococcal strains in a human population — but only because we knew in advance to focus on the genes targeted by the vaccine<sup>3</sup>. Our ability to identify functional differences in closely related genes is rarely sophisticated enough to pull out important genes or networks if we do not know what to look for in the first place. Moreover, genomes are littered with clues both true and false, such as 'hypothetical proteins' and genes that are understood poorly or not at all, but could make for important differences in what metabolic networks do.

We need to be able to identify functional differences in closely related genes from sequence alone. Until then, we must remember that apparent similarities might cloak important differences.

**Does the study show causation or just correlation?** A separate question is raised when distinct microbiomes can be identified and associated with diseases or other conditions. Then we are left with the chestnut of causes and correlates. Sometimes, a particular microbiome found in association with disease will be merely a bystander<sup>4</sup>.

A 2012 article comparing the gut microbiomes of old people living in care homes with those of old people living in the community found distinct microbiomes that correlated with multiple scores of frailty<sup>5</sup>. After accounting for some potentially con-

*"Press officers must stop exaggerating results, and journalists must stop swallowing them whole."*

founding factors, the authors proposed a causal relationship: diet altered the microbiome, which in turn altered health. This explanation fits the data, but the reverse causality — the potential for poor health to alter the gut microbiome — was not explored. Frailer people probably have less active immune systems and differences in digestion (such as the time required for food to pass through the stomach and intestines) — factors that could change the microbiome. This work is not the only example of this sort of confusion.

**What is the mechanism?** All scientists are taught the catechism that correlation is not causation, but correlation almost always implies some sort of causal relationship. We just don't know what it is. We must determine it with careful experiments.

In the past three or four years, studies have advanced from characterizing a broad community of mainly unculturable microbes to identifying functional elements, individual taxa or particular properties. We can now design experiments to precisely define actions of components of the microbiome<sup>6</sup>, for example by reconstituting communities but leaving out specific taxa, or by precisely measuring the biochemical activity of an experimental microbiome in an 'organ on a chip'<sup>7</sup>. A return to a reductionist approach is essential if we are to pinpoint both whether the microbiome affects human health, and exactly how it does so.

### How much do experiments reflect reality?

Even if the microbiome can have an experimental effect, it may not be an important cause of the symptoms seen in ill people.

Much work has addressed the role that gut flora have in obesity, and several studies have found associations between the gut microbiome and weight gain<sup>8</sup>. To assess whether this association was cause or consequence, researchers collected gut-microbiome samples from human twins (one obese, one not)

and introduced the microbiota to mice. Mice previously colonized with an 'obese' microbiome lost weight when supplied with a 'lean microbiome', but only if also fed a normal or low-fat diet. Diet alone had little effect<sup>9</sup>. Although this elegantly controlled experiment suggests great potential for the microbiome and related therapies to affect health, it also shows the microbiome's limits: the effect was dependent on other factors, in this case diet.

Microbiome studies often rely on germ-free mice. These animals allow researchers to readily introduce an experimental microbiota. But they do not represent the animals' natural state and are typically unhealthy owing to the lack of a microbiome. So results may not predict responses in animals with flourishing microbiomes. Mice and their microbiomes are also adapted to a rather different niche from humans, so results may not be generalizable.

### Could anything else explain the results?

There are good reasons to think that bacteria influence us in a host of ways. But there are many other — possibly more important — influences, such as diet in the earlier example. Whenever a study links a microbiome to a disease, wise critics should ask whether other contributors to disease are considered, compared and reported.

The hype surrounding microbiome research is dangerous, for individuals who might make ill-informed decisions, and for the scientific enterprise, which needs to develop better experimental methods to generate hypotheses and evaluate conclusions. Funding agencies must not let their priorities be distorted by the buzz around the field, but look dispassionately at the data. Press officers must stop exaggerating results, and journalists must stop swallowing them whole. In pre-scientific times when something happened that people did not understand, they blamed it on spirits. We must resist the urge to transform our microbial passengers into modern-day phantoms. ■

**William P. Hanage** is associate professor of epidemiology at the Harvard School of Public Health in Boston, Massachusetts, USA.  
e-mail: whanage@hsph.harvard.edu

1. Blaser, M. J. *Missing Microbes: How the Overuse of Antibiotics is Fueling our Modern Plagues* (Henry Holt and Co., 2014).
2. Wilkins, M. R. et al. *Proteomics* **6**, 4–8 (2006).
3. Croucher, N. J. et al. *Nature Genet.* **45**, 656–663 (2013).
4. Shanahan, F. *Nutr. Rev.* **70** (suppl. 1), S31–S37 (2012).
5. Claesson, M. J. et al. *Nature* **488**, 178–184 (2012).
6. Goodman, A. L. et al. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).
7. Huh, D. et al. *Nature Protocols* **8**, 2135–2157 (2013).
8. Ley, R. E. *Curr. Opin. Gastroenterol.* **26**, 5–11 (2010).
9. Ridaura, V. K. et al. *Science* **341**, 1241214 (2013).





Aristotle (384–322 BC) was inspired by observation of the natural world on the Greek island of Lesbos.

## HISTORY OF SCIENCE

# The first scientist

**Roberto Lo Presti** applauds a brilliant reappraisal of Aristotle as the father of observational biology.

Aristotle is considered by many to be the first scientist, although the term postdates him by more than two millennia. In Greece in the fourth century BC, he pioneered the techniques of logic, observation, inquiry and demonstration. These would shape Western philosophical and scientific culture through the Middle Ages and the early modern era, and would influence some aspects of the natural sciences even up to the eighteenth century.

Armand Marie Leroi's reappraisal of this colossus, *The Lagoon*, is one of the

most inspired and inspiring I have read. It combines a serious, accessible overview of Aristotle's methods, ideas, mistakes and influence with a contextualizing travelogue that also found expression in Leroi's 2010 BBC television documentary *Aristotle's Lagoon*. Leroi's ambitious aim is to return Aristotle to the pantheon of biology's greats, alongside Charles Darwin and Carl Linnaeus. He has achieved it.

Leroi, an evolutionary developmental biologist, visits the Greek island of Lesbos — where Aristotle made observations

of natural phenomena and anatomical structures — and puts his own observations in dialogue with those of the philosopher. It was in the island's lagoon of Kolpos Kalloni that Aristotle was struck by the anatomy of fish and molluscs, and started trying to account for the function of their parts. Leroi's vivid descriptions of the elements that inspired Aristotle's biological doctrines — places, colours, smells, marine landscapes and animals, and local lore — enjoin the reader to grasp them viscerally as well as intellectually.

Aristotle's time on Lesbos was only a chapter in a life of discoveries, and Leroi covers those signal achievements with breadth and depth. He details the theoretical and methodological principles governing the functional anatomy of species from pigeons to tortoises, discussed by Aristotle in *On the Parts of Animals*, as well as the descriptive zoology expounded in his *History of Animals*. For instance, Leroi explores Aristotle's theory of causation, based on the distinction between material, efficient, formal and final causes. He looks at the philosopher's views on the directedness of natural phenomena and the role played by necessity and hazard. He sketches out the theory of four elements (fire, air, water and earth) as the prime constituents of natural bodies. And he looks at the theory of soul and its relationship to the body — through which Aristotle accounted for aspects of physiology and psychology, from nutrition to rational thinking.

Fascinating chapters are devoted to Aristotle's gradualist conception of the natural world and living things — perhaps best expressed in the saying *natura non facit saltum*, or 'nature does not make jumps'. Also covered is his theory of sexual generation and transmission of hereditary traits, which he expounded in the masterful *On the Generation of Animals*. Despite a number of mistaken assumptions (such as the lack of a female 'seed'), this theory encompasses a huge number of valuable observations and insights that laid the foundations of modern embryology.

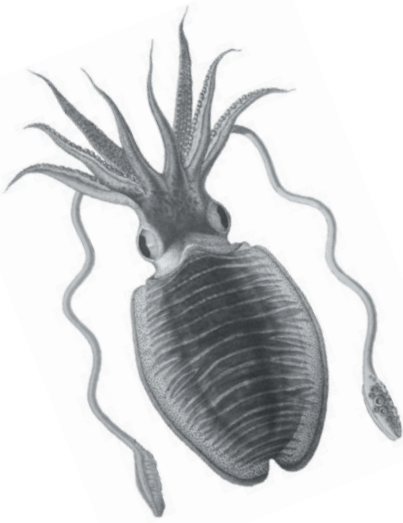
*The Lagoon* traces other ways in which Aristotelian thought has permeated West-

ern science. Leroi charts its influence on Renaissance anatomists and physiologists. The English physician William Harvey's discovery of blood circulation, for instance, was largely inspired by Aristotle's biological ideas, especially the concept of the heart as the most important organ in the body, as well as by



**The Lagoon: How Aristotle Invented Science**

ARMAND MARIE LEROI  
Bloomsbury/Viking:  
2014.



The cuttlefish was among many creatures that Aristotle observed.

BLOOMSBURY

Aristotle's empirical emphasis on investigation and demonstration. Leroi shows how masters of comparative anatomy including Georges Cuvier (1769–1832) took inspiration from Aristotle in describing and comparing the parts of animals in light of their function as well as of their shape. He compares Aristotle's theories with the thinking of taxonomists such as Linnaeus, of Darwin on evolution, and of the twentieth-century fathers of systems theory and cybernetics such as Walter Cannon and Norbert Wiener.

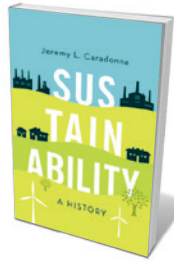
Leroi is careful not to represent Aristotle as a precursor in crude terms, or to read him through inappropriate contemporary lenses. Instead, he highlights aspects of Aristotle's doctrines that still 'speak' to contemporary scientists, and that have been illuminated by modern scientific understanding — for example, Aristotle's emphasis on direct observation and dissection. The philosopher argued, Leroi explains, "that 'knowing' in the sense of 'perceiving' is the foundation of 'knowing' in the sense of 'understanding'".

As Leroi acknowledges, decades of scholarly effort by philosophers and historians such as Allan Gotthelf and James Lennox have gone into the reassessment of Aristotelian biology and its effect on the history of Western science. In this respect, the book broaches no new questions, and brings no new perspective to the heated debates among Aristotelian scholars.

But that is to miss its point. *The Lagoon* is a wonderful introduction to Aristotle's biology, which specialists will also enjoy. Every page is a reminder of the great beauty that we can experience by seeing the world through Aristotelian eyes. ■

**Roberto Lo Presti** lectures on classics and ancient philosophy and coordinates the doctoral programme on the history of ancient science at the Berlin Graduate School of Ancient Studies, part of the Humboldt University of Berlin.  
e-mail: roberto.lo.presti@hu-berlin.de

## Books in brief



### Sustainability: A History

Jeremy L. Caradonna OXFORD UNIVERSITY PRESS (2014)

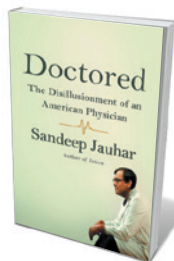
As a concept, sustainability is now near-ubiquitous. But is it a "buzzless buzzword", as environmentalist Bill McKibben has opined? Historian Jeremy Caradonna writes that, on the contrary, this dynamic ethos has plenty of buzz. Predicated on joined-up thinking (such as the idea that society, economy and environment are linked), it emerged with seventeenth-century concerns over European deforestation and is now, Caradonna posits, the keystone of solutions to looming global crises. An exemplary study of an idea's long march through domains from urbanism to social justice.



### The Marketplace of Attention: How Audiences Take Shape in a Digital Age

James G. Webster MIT PRESS (2014)

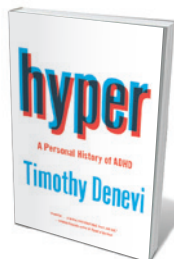
The legions of social networks, news outlets and other digital media all jostle for a limited resource — human attention. As James G. Webster notes in this uneven but fascinating study, "It's a zero sum game that dooms most offerings to obscurity". His investigation of what pulls audiences in trounces prevailing thinking, such as the theory that audiences are zombies herded into "filter bubbles" by data-driven choices. Instead, he reveals a "massively overlapping culture" in which commonality remains surprisingly high.



### Doctored: The Disillusionment of an American Physician

Sandeep Jauhar FARRAR, STRAUS AND GIROUX (2014)

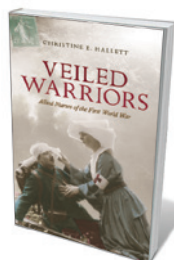
In this absorbing memoir-cum-analysis, Sandeep Jauhar traces his years as a fledgling cardiologist against the backdrop of a health-care system in peril. US medicine emerges as an arena in which the physicians suffer as much as the patients: out of 12,000 doctors surveyed in 2008, Jauhar notes, just 6% reported positive morale in their colleagues. Factors such as vast medical-school debts, grinding overwork and the rise of autonomy-eroding health-maintenance organizations are leaving many medics reeling and many potential wannabes seeking other fields. An impassioned call to action.



### Hyper: A Personal History of ADHD

Timothy Denevi SIMON AND SCHUSTER (2014)

It began with screaming in a world reduced to a colourless blur. Writer Timothy Denevi was diagnosed with attention deficit hyperactivity disorder (ADHD) at the age of six, and in this haunting narrative he explores the world's most scrutinized childhood condition from the inside out — a litany of school conflicts and rounds of evolving treatments. Through it, Denevi interweaves ADHD's historical trajectory and recent findings, from difficulties with diagnosis (the symptoms are easily conflated with 'normal' childhood behaviour) to the brain regions implicated. Denevi has survived, but at a cost.



### Veiled Warriors: Allied Nurses of the First World War

Christine E. Hallett OXFORD UNIVERSITY PRESS (2014)

Professional nursing during the Great War is often seen through the gauze of romantic myth-making, notes historian Christine E. Hallett in this stinging chronicle. In fact, the trained nurses of the Allied forces were less noble young helpmeets than a heterogeneous group of tough-minded women. Eager for formal social and political recognition, they were also faced with grisly new medical challenges such as gas gangrene. As Hallett writes, they fought "a multi-layered battle: for lives, for recognition, and for equality". **Barbara Kiser**



# Second rock from the Sun

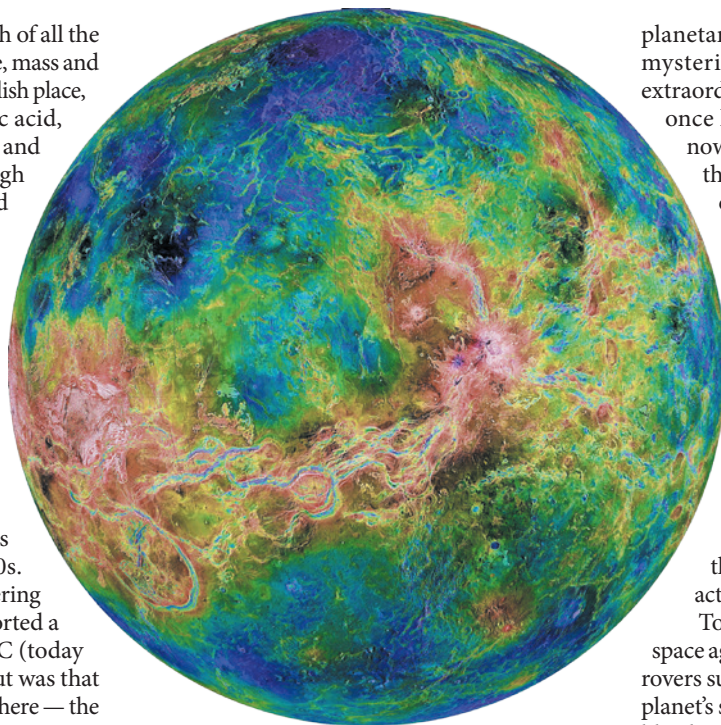
Andrew P. Ingersoll relishes a study of scientific discoveries on hot, toxic Venus.

Venus is the closest to Earth of all the planets — in distance, size, mass and composition. But it is a hellish place, with opaque clouds of sulphuric acid, crushing atmospheric pressures and surface temperatures high enough to melt lead. Not photogenic and too blistering for conventional rovers and landers, Venus has been passed over by the big missions. Nevertheless, Venus exploration and the interpretation of the resulting data have had their fair share of exciting moments. It is this history that physicist Fredric Taylor relates in his informative primer, *The Scientific Exploration of Venus*.

The first hints that Venus was not like Earth came in the 1950s. Earth-based radio telescopes, peering through the planet's clouds, reported a temperature of more than 300 °C (today the estimate is rather higher). But was that the temperature of Venus's ionosphere — the upper atmosphere — or its surface?

Astronomer Carl Sagan, whose doctoral thesis showed how the greenhouse effect could account for high surface temperatures on Venus, set out to test his theory. He lobbied the newly established NASA to build the first interplanetary spacecraft and equip it with the right instrument — a microwave radiometer. Identical spacecraft, Mariner 1 and 2, were built and launched in 1962. Mariner 1 failed, but Mariner 2 made the three-month trip to Venus and took the crucial measurement, showing that the temperature was greatest when the radiometer looked straight down at the surface and least when it looked horizontally through the ionosphere.

Taylor also tells the story of the US–Soviet race to explore Venus. Relative to the cold war, it was almost friendly, and indeed may have helped the slow easing of tensions between the two superpowers. Scientists read each others' papers and started to visit each others' countries. Mutual respect began to emerge, but it was not a smooth process. In 1967, the Soviet spacecraft Venera 4 parachuted into Venus's atmosphere. It measured a pressure of 2,000 kilopascals and a temperature of 275 °C before it stopped transmitting. Soviet scientists insisted that the probe had reached the planet's surface, although there was already evidence that temperatures and pressures there would be higher. A newly minted PhD at the time, I remember being struck by their



Venus, seen here in a composite radar image, is Earth's closest planetary neighbour.

insistence, thinking that perhaps they were not allowed to admit failure.

But the Soviets persisted, and in 1969, Venera 5 and 6 made the first scientifically successful landings on the Venusian surface. Once down, they measured temperature, pressure and composition, surviving for an hour before succumbing to the searing heat. Now it was clear that the surface was very hot, 450 °C, and the pressure was 9,000 kilopascals — 90 times that of Earth's atmosphere. In 1975, Venera 9 and 10 took images of the surface, revealing flat, cracked plates of rock that could only be solidified lava. The pictures were taken by a television camera in natural light, indicating that about 2% of the incident sunlight was getting through the massive atmosphere to the ground. "Like Moscow on a cloudy winter day," said one of the Soviet scientists at a conference I attended.

Taylor also reviews the origin, evolution and composition of

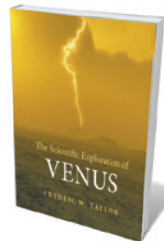
planetary bodies, and Venus's remaining mysteries. What maintains the planet's extraordinary greenhouse effect? If Venus once had an ocean, where is the water now? Why does the wind at the top of the clouds blow 50 times faster than on the ground, at more than twice the speed of Earth's jet streams? And why does Venus have an atmosphere of carbon dioxide almost 100 times more massive than Earth's nitrogen–oxygen atmosphere? For Venus's surface and interior, the big questions are: how does the planet shed its internal heat, given that there is no sign of plate tectonics? Are volcanoes enough? Are the volcanoes active and what gases do they bring to the surface? Is volcanic activity changing the climate of Venus?

To answer some of these questions, space agencies are studying aeroplanes and rovers suitable for Venus. Operating on the planet's surface is difficult but not impossible: the instruments must either operate at high temperature, or be refrigerated.

The book reproduces numerous intriguing images of the surface taken by the radar on NASA's Magellan spacecraft. A chapter on the clouds, which are concentrated at 50–65 kilometres altitude (Earth's are at 0–15 kilometres) reveals much about the meteorology of the planet. Despite the stronger winds, the circulation of the atmosphere is smoother than Earth's — less interrupted by waves and eddies, with winds blowing more nearly east to west. Also, the massive atmosphere seems to have evened out the temperature difference between equator and pole. At the poles, the clouds mark the edges of a vortex that resembles the polar vortex on Earth.

Taylor clearly believes that Venus has a lot to tell us about our own planet — particularly through its greenhouse effect, winds and volcanoes, and the history of its surface. He writes wistfully of "the forgotten world" and the "unfulfilled objectives" of never-flown missions. The last part of the book is, however, forward-looking, covering plans and visions for further exploration of Earth's wayward next-door neighbour. ■

Andrew P. Ingersoll is professor of planetary science at the California Institute of Technology in Pasadena. He is the author of *Planetary Climates*.  
e-mail: [api@gps.caltech.edu](mailto:api@gps.caltech.edu)



**The Scientific Exploration of Venus**  
FREDRIC W. TAYLOR  
Cambridge University Press: 2014.

NASA/JPL/USGS

# Correspondence

## UK bill could prompt biodiversity loss

The UK government's proposed Infrastructure Bill for England and Wales gives new powers to control or eradicate invasive, non-native species (see [go.nature.com/kbkvtt](http://go.nature.com/kbkvtt)). However, what constitutes such a species needs careful definition to ensure that any use of these powers is beneficial for conservation.

The draft bill defines non-native species as those that are "not ordinarily resident in, or a regular visitor to, Great Britain". This definition covers past native species that are now extinct, species that may become naturally established under a changing climate, and species listed in Schedule 9 of the Wildlife and Countryside Act. That list contains native species, including some that have been reintroduced into the wild, such as the barn owl (*Tyto alba*), capercaillie (*Tetrao urogallus*), chough (*Pyrrhocorax pyrrhocorax*) and red kite (*Milvus milvus*). The legislation could also preclude future species reintroductions, a tool to counter biodiversity loss.

The current definition has serious implications for wildlife management. Once a species is classified as non-native, it can also be classified as invasive — and would therefore be subject to invasive-species legislation.

Proposed amendments to address these problems have been rejected. If the bill is passed in its present form, it could lead to an irreversible loss of native biodiversity.

**Sarah Durant\*** *Institute of Zoology, Zoological Society of London, UK.*  
[sarah.durant@ioz.ac.uk](mailto:sarah.durant@ioz.ac.uk)

*\*On behalf of 24 correspondents (see [go.nature.com/9eembe](http://go.nature.com/9eembe) for full list).*

## Don't let microbial samples perish

Microbial ecologists must coordinate to archive sample collections and genetic material. This will prevent valuable

specimens from being lost to science and allow rigorous assessment of the effects of globally changing factors, disease and pollution on microbial communities.

Archiving is particularly valuable for hard-to-obtain or irreplaceable samples: for example, those from deep-sea hydrothermal vents or subglacial lakes in Antarctica. Samples may form an important time series, as in faecal material from extant populations never exposed to antibiotics, or soils from high-latitude systems threatened with biodiversity loss because of climate change.

Data generation is cheap, and getting cheaper. Technologies now exist to store DNA at room temperature for long periods and to reanalyse samples directly, which is better than trying to cobble together previous data sets generated using outdated methods. Reanalysis can also enable comparison of samples collected at different sites or time points.

In microbial ecology, proper sample archiving will accelerate advances, as collections of plants, animals and cultures have done for other areas of biology.

**Noah Fierer** *University of Colorado, Boulder, USA.*

**Craig Cary** *University of Waikato, Hamilton, New Zealand.*  
[caryc@waikato.ac.nz](mailto:caryc@waikato.ac.nz)

## Risk review is under way for invasive toad

Sven Mecke and colleagues call for prior assessment of ecological risks that might be associated with eradication measures against the invasive Asian common toad *Duttaphrynus melanostictus* (*Nature* **511**, 534; 2014). The Amphibian Specialist Group in Madagascar — part of the International Union for Conservation of Nature's Species Survival Commission network — is already undertaking such an evaluation, along with local and international experts.

We disagree that the effects of an invasive species need to be fully understood before starting control operations. Experience with other invaders shows that this could take decades, and swift action is crucial to stop an invasion from becoming widespread. We already know that *D. melanostictus* is invasive elsewhere in the tropics and is a biosecurity hazard in Australia.

Nationally coordinated by Christian Randrianantoandro, our efforts include determining the toads' distribution, providing educational materials to local communities and assembling experts to develop the feasibility study. We shall use genetic analyses to identify the source of introduction and will screen toads for pathogens and parasites. All toad sightings in Madagascar have so far been in urban and nearby degraded habitats, which would limit any threat to native biota should an eradication programme be carried out.

Costs of the preliminary assessment are estimated at US\$50,000. The Amphibian Survival Alliance is running an online fund-raising campaign, and we hope to involve international non-governmental organizations. A globally coordinated response may still stop the toads from invading Madagascar.

**Franco Andreone\*** *Museo Regionale di Scienze Naturali, Turin, Italy.*

[franco.andreone@gmail.com](mailto:franco.andreone@gmail.com)

*\*On behalf of 11 correspondents (see [go.nature.com/cdpbbq](http://go.nature.com/cdpbbq) for full list).*

## Ageing: develop models of frailty

Good preclinical models of ageing are needed to discover the molecular mechanisms behind declining human physical performance (*Nature* **511**, 405–407; 2014). The latest animal models of frailty are a step in the right direction.

For example, a genetically modified frail-mouse model

mimics the inflammation and weakness that often afflicts older people (see A. Akki *et al.* *Age* **36**, 21–30; 2014). Frailty can also be modelled in naturally ageing mice as a frailty-phenotype score, graded by such performance measures as grip strength and walking speed (H. Liu *et al.* *J. Gerontol. A* <http://doi.org/t6p>; 2013).

We and others have quantified a clinical frailty index in mice by tracking the accumulation of age-related deficits. This index increases with age in the same way as it does in humans (J. C. Whitehead *et al.* *J. Gerontol. A* **69**, 621–632; 2014).

More-sophisticated animal models of frailty will need to include a broader range of performance measures if they are to properly represent the condition in people.

**Susan E. Howlett, Kenneth Rockwood** *Dalhousie University, Halifax, Nova Scotia, Canada; and University of Manchester, UK.*  
[susan.howlett@dal.ca](mailto:susan.howlett@dal.ca)

## Ageing: research needs social science

Translational biomedical research into ageing and longevity needs to include social science if it is to produce interventions for slowing physiological decline (*Nature* **511**, 405–407; 2014).

Promoting a healthy lifespan depends on social factors as well as on medical insight. Any study on caloric restriction, for instance, should consider the cultural background of participants.

Social science also needs to be incorporated into therapeutic investigations — for example, to understand why some individuals do not take their medication. Such insight would boost compliance and therefore drug effectiveness.

**Philippe de Souto Barreto** *Institute of Ageing, University Hospital of Toulouse (CHU Toulouse), France.*  
[philipebarreto81@yahoo.com.br](mailto:philipebarreto81@yahoo.com.br)



## Microbes eat rock under ice

The first description of the microorganisms inhabiting a subglacial lake deep below the Antarctic ice sheet reveals some of the complex interactive metabolic processes that sustain these microbial communities. [SEE LETTER P.310](#)

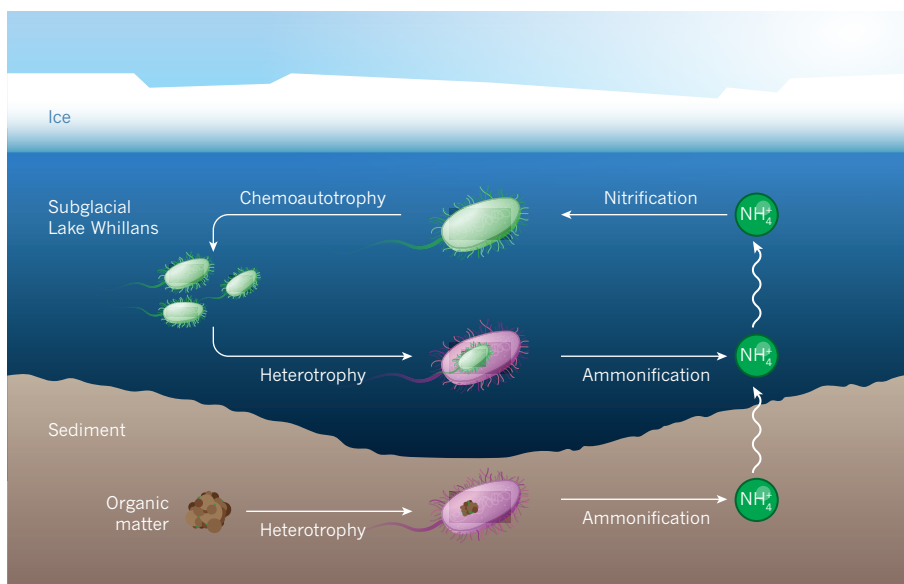
MARTYN TRANTER

The deep cold marine biosphere<sup>1</sup> has a newly described and highly active fresh-water neighbour. On page 310 of this issue, the WISSARD Science Team and other authors (Christner *et al.*<sup>2</sup>) show that the cold fresh waters under the thick ice of the West Antarctic Ice Sheet are a habitat for microbial life, joining communities already found in deep cold ocean waters and sediments. The report is a landmark for the polar sciences, demonstrating unequivocally for the first time by direct, contamination-free sampling that microbes are present in the waters and sediment of Subglacial Lake Whillans, located some 800 metres below the surface of the ice and at a temperature of  $-0.17^{\circ}\text{C}$ .

The discovery of microorganisms in this environment adds to our appreciation that glaciers and ice-sheet beds are not sterile, but in fact host diverse microbial communities<sup>3</sup>. The beds of small, valley glaciers were shown to contain microbial communities only 15 years ago<sup>4</sup>. However, these glaciers have much shallower ice than the West Antarctic Ice Sheet, and microorganisms, organic matter and chemical species (including nutrients such as dissolved oxygen and nitrate) are exported from their melting ice surfaces down to the bed, so it is easy to see how microbes are able to grow there.

This is not the case for the West Antarctic Ice Sheet, where there is little surface melt away from the ice margins, and even less likelihood that this water can find a way through a kilometre or so of ice at temperatures well below freezing. Instead, water is produced from geothermal heating at the bed and through frictional melting during ice flow<sup>5</sup>. This means that any microorganisms living in the water that is present beneath about 55% of Antarctica<sup>5</sup> must exist on energy and nutrient sources that come from melting ice, rock and sediment beneath the ice, and from recycling of materials from dead microorganisms<sup>6</sup>. The team's results give insight into how this may happen.

Microbes living in the deep ocean can rely on the remains of surface organisms raining down from above as energy sources. By contrast, microbes in the deep cold freshwater environment of Subglacial Lake Whillans



**Figure 1 | Around goes the ammonium ion.** Christner *et al.*<sup>2</sup> have shown that significant cycling of ammonium ions ( $\text{NH}_4^+$ ) occurs in Subglacial Lake Whillans, 800 metres beneath the surface of the West Antarctic Ice Sheet. They show that this cycling is conducted by chemoautotrophs (bacteria that use ammonium ions instead of sunlight to drive their vital processes and proliferation) and heterotrophs (bacteria that decompose organic matter, including other bacteria, to derive energy, producing ammonium ions as a consequence). The chemoautotrophs are probably confined to the lake's water column, whereas the heterotrophs probably exist in both the water column and in the lake's sediment.

must use energy sources contained in minerals crushed from bedrock by the ice, including sulphides (for example pyrite, found in many rock types) and reduced iron,  $\text{Fe(II)}$ , which is found in many minerals containing iron and magnesium (olivines, pyroxenes and micas, for instance). These reduced iron and sulphur compounds, along with dead microorganisms, can be oxidized by oxygen in the water. This liberates the energy necessary to drive the vital processes of the living microorganisms that enable or catalyse the mineral-oxidation reactions<sup>7</sup>. In this sense, the microbes 'eat rock' — although in practice, they attach to the mineral particles and help them to dissolve.

The microorganisms needed to promote these types of reaction have been found by the team, in particular Proteobacteria, which make up around 53% of the gene sequences examined. Glacier-crushed sediment is a ready source of phosphorus<sup>8</sup>, a key nutrient for enabling microbial growth, but sources of the other key nutrient — dissolved nitrogen

species such as nitrate ( $\text{NO}_3^-$ ) and ammonium ( $\text{NH}_4^+$ ) — are harder to come by. They include a little from ice that melted to produce the water<sup>7</sup> and small amounts in some micas and feldspars<sup>6</sup>. Any nitrogen scavenged from rock is hard-won, so must be preserved or tightly recycled to keep the microbial ecosystems viable (Fig. 1).

An intriguing feature of the team's data is that ammonium, often liberated from the decomposition of organic matter, is the principal dissolved inorganic nitrogen species. Furthermore, the water column contains significant numbers of nitrifying microorganisms, which oxidize ammonium to nitrite and nitrate. The particular ratio of oxygen isotopes in the nitrate —  $\Delta^{17}\text{O}$  of  $\text{NO}_3^-$  — is almost zero, and is consistent with nitrification. This strongly suggests that hard-won ammonium diffusing up from the sediment is tightly recycled between dead and living microorganisms, and is a means of sustaining new microbial growth — a process known as chemoautotrophy.

The authors also find that surface sediments beneath the water contain organic matter with higher carbon-to-nitrogen ratios than the organic matter in the water column. This also suggests that decomposition of organic matter in the sediments releases ammonium and other dissolved nitrogen species back into the water column. Some of the carbon in the decomposing organic matter seems to be released as acetate and formate — forms of dissolved organic carbon that can readily be taken up by microbes in the water column. Hence, processes in the sediment, and recycling of materials between the sediment and the water column, are important for the longevity of the microbial ecosystem in the lake.

In this respect, this deep cold freshwater ecosystem is similar to those in and overlying deep cold ocean sediments. The West Antarctic

Ice Sheet is draped over deep basins of former marine sediments, kilometres thick, which also contain organic matter<sup>9</sup>. Just how much influence this gradually decomposing organic matter has on surface sediments and the waters between the sediment and the ice bed remains to be seen, but beneath some parts of Antarctica, where water flow is slow, one can easily imagine how chemicals diffusing up from deeper sediment might allow microbial communities to exist in the shallower sediments (Fig. 1).

The team has opened a tantalizing window on microbial communities in the bed of the West Antarctic Ice Sheet, and on how they are maintained and self-organize. The authors' findings even beg the question of whether microbes could eat rock beneath ice sheets on extraterrestrial bodies such as Mars<sup>10</sup>. This idea has more traction now. ■

**Martyn Tranter** is at the Bristol Glaciology Centre, School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK. e-mail: m.tranter@bristol.ac.uk

1. Parkes, R. J. *et al.* *Nature* **436**, 390–394 (2005).
2. Christner, B. C. *et al.* *Nature* **512**, 310–313 (2014).
3. Anesio, A. M. & Laybourn-Parry, J. *Trends Ecol. Evol.* **27**, 219–225 (2012).
4. Sharp, M. *et al.* *Geology* **27**, 107–110 (1999).
5. Pattyn, F. *Earth Planet. Sci. Lett.* **295**, 451–461 (2010).
6. Tranter, M. & Wadham, J. L. in *Treatise on Geochemistry* 2nd edn, Vol. 7 *Surface and Ground Water, Weathering, and Soils* (ed. Drever, J. I.) 157–173 (Elsevier, 2014).
7. Christner, B. C. *et al.* *Limnol. Oceanogr.* **51**, 2485–2501 (2006).
8. Hodson, A., Mumford, P. & Lister, D. *Hydrol. Processes* **18**, 2409–2422 (2004).
9. Wadham, J. L. *et al.* *Nature* **488**, 633–637 (2012).
10. Cockell, C. S. *et al.* in *Antarctic Subglacial Aquatic Environments* Vol. 192 (eds Siegert, M. J. & Kennicutt, M. C.) 129–148 (Am. Geophys. Union, 2013).

## DEVELOPMENTAL BIOLOGY

# It takes muscle to make blood cells

**Blood stem cells derive at least in part from an embryonic vessel called the dorsal aorta. It emerges that a flanking tissue called the somite contributes cells and signals to this process. SEE LETTERS P.314 & P.319**

SUPHANSA SAWAMIPHAK &  
DIDIER Y. R. STAINIER

The extrinsic cues that instruct cells to become blood-cell precursors are mostly unknown. Studies in different vertebrate models have shown that these precursors, called haematopoietic stem cells, originate at least in part from the first functional intra-embryonic blood vessel, the dorsal aorta. However, the cell types and signalling molecules that promote the generation of haematopoietic stem cells in the dorsal aorta are not well understood. Two papers in this issue<sup>1,2</sup> find that structures called somites, precursors of a range of tissues including the vertebrae and skeletal muscle, are involved at more than one stage of this developmental process.

Haematopoietic stem cells (HSCs) have the capacity to replenish all blood-cell types throughout life. Live-imaging experiments in zebrafish (*Danio rerio*)<sup>3–5</sup> provided the first conclusive evidence that these cells derive from the endothelial cells that line the dorsal aorta (DA). However, these and other studies<sup>6</sup> indicated that only a subset of DA endothelial cells can become HSCs.

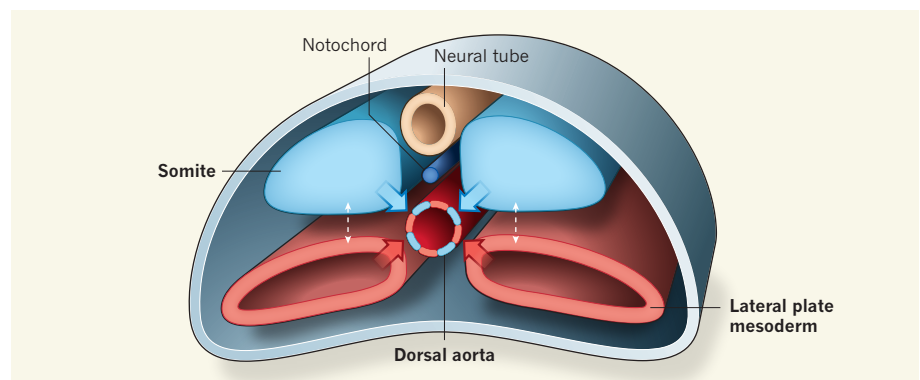
Studies in avian embryos<sup>7</sup> gave the first indication that the ability of endothelial cells to become HSCs was determined by their

origin. Endothelial cells from a tissue called the lateral plate mesoderm populate part of the DA, and can give rise to HSCs<sup>7</sup>. Endothelial cells from the somites populate another part, and do not become HSCs<sup>8</sup>. When production of HSCs within the DA ceases, the somite-derived endothelial cells replace those from the lateral plate mesoderm<sup>8</sup>. It therefore stands to reason that somite-derived cells and signalling

factors may regulate the endothelium-to-HSC transition.

Nguyen *et al.*<sup>1</sup> (page 314) investigated the role of the somites in development of the DA and HSCs in zebrafish using a mutant fish strain called *choker*, in which signalling from somitic tissues is defective<sup>9</sup>. They expanded on the previous analysis of this defect, and found that *choker* fish harbour a defective copy of *meox1*, a gene that is normally expressed in the somites. They then examined HSC development in these mutants. Surprisingly, these animals had more HSCs than their wild-type siblings.

Next, the authors used genetic tools to trace the descendants of cells within the somites (an approach known as lineage tracing). They found that, as in birds and mice<sup>7,8</sup>, a portion of zebrafish somitic cells give rise to endothelial cells that line the DA — a progenitor population that Nguyen and colleagues named endotomal cells. Up- and downregulating *meox1* function revealed that the gene inhibits the formation



**Figure 1 | The birth of blood-cell precursors.** A simplified cross-section through a zebrafish embryo illustrates how haematopoietic stem cells (HSCs) arise in the dorsal aorta (DA), an embryonic structure lined by endothelial cells from two different origins. Nguyen *et al.*<sup>1</sup> report that endothelial cells that migrate to the dorsal aorta from the somites (blue arrows) do not give rise to HSCs themselves, but instead help other endothelial cells in the DA to become HSCs. Kobayashi *et al.*<sup>2</sup> find that endothelial precursors that migrate from the lateral plate mesoderm (red arrows) interact with the developing somite en route to the DA (dashed arrows). Some of these cells go on to become HSCs.



of endotomal cells in the somites. Moreover, the authors found that endotomal cells do not give rise to HSCs themselves, but help other endothelial cells to become HSCs (Fig. 1).

Accordingly, ablation of endotomal-derived endothelial cells led to a significant reduction in the number of HSCs. This outcome was, at least in part, due to reduction of the signalling protein Cxcl12b. When Nguyen and co-workers either chemically inhibited or 'knocked down' Cxcl12b function, they saw a drastic reduction in HSC numbers.

Given the small size, elongated shape and tight packing of endothelial cells and their non-endothelial neighbours, it will be important to follow up this and other lineage-tracing studies<sup>10,11</sup> in mice, for example by using an approach that marks somite-derived cells only if they become endothelial cells. One interpretation of the studies in birds<sup>8</sup> is that replacing the endothelial cells derived from the lateral plate mesoderm with somite-derived endothelial cells limits the DA's ability to generate HSCs. In light of Nguyen and colleagues' evidence that reducing the number of somite-derived endothelial cells actually decreases HSC formation in zebrafish, it will be essential to directly test the role of somite-derived endothelial cells in HSC emergence in mice.

Do the somites also regulate HSC precursors at earlier stages? The lateral plate mesoderm is in direct contact with somitic tissue, and there is evidence<sup>12</sup> that somite-derived signalling proteins promote HSC formation as cells migrate from the lateral plate mesoderm to the forming DA. Wnt16 may be one such protein<sup>12</sup>, although Nguyen and co-workers question this finding, because *wnt16* expression is in fact downregulated in *choker* mutants.

In addition to secreted factors, direct physical interactions between migrating cells and somitic tissues might regulate HSC generation. Kobayashi *et al.*<sup>2</sup> (page 319) report that cell-adhesion proteins called junction adhesion molecules (Jams) are involved in interactions between HSC precursors and somites, which in turn are required for HSC formation. They found that endothelial precursors in the lateral plate mesoderm express *jam1a* as they migrate towards the forming DA along the somite surface, which expresses *jam2a*. The two Jam proteins physically interact, promoting strong cell-cell contacts.

Kobayashi and colleagues also found that the somites express the genes *deltaC* and *deltaD*, which encode Notch-binding proteins, and that the migrating endothelial precursors express Notch protein. Notch signalling is involved in cell-cell communication in many settings in the body. Knockdown of *jam1a* led to loss of Notch signalling and loss of HSCs, but HSC formation could be restored in these embryos with forced activation of Notch signalling in endothelial precursors. Thus, strong Jam-mediated cell-cell interactions might facilitate the activation of Notch

signalling in endothelial precursors.

This work suggests that Notch signalling is required in HSC progenitors during their migration from the lateral plate mesoderm, earlier than previously reported<sup>12</sup>. This model has implications for optimizing the induction of HSCs *in vitro*, which remains challenging, probably because some key factors are missing from the protocols being used.

In addition to Kobayashi and co-workers' finding, another study<sup>13</sup> has shown that somite-derived endothelial cells themselves require endogenous Notch signalling. To analyse the role of Notch signalling in more detail, tools that permit precise temporal and spatial control of gene expression will therefore be needed. These tools will include conditional genetic mutants that can be induced to lack gene function only at certain times and in certain tissues, technology that for the most part is not yet available in zebrafish.

Together, these two studies show that somites play a key part in HSC formation. The idea that cells from the somites populate embryonic blood vessels is now fairly well established, but their role in HSC formation in birds and mammals remains to be defined. Ultimately, of course, understanding where

the various cell types come from should help to determine the identity and exact sequence of signalling pathways activated in HSCs, their precursors and their derivatives. ■

**Suphansa Sawamiphak and Didier**

**Y. R. Stainier** are in the Department of Developmental Genetics, Max Planck Institute for Heart and Lung Research, Bad Nauheim 61231, Germany.

e-mails: [suphansa.sawamiphak@mpi-bn.mpg.de](mailto:suphansa.sawamiphak@mpi-bn.mpg.de); [didier.stainier@mpi-bn.mpg.de](mailto:didier.stainier@mpi-bn.mpg.de)

1. Nguyen, P. D. *et al.* *Nature* **512**, 314–318 (2014).
2. Kobayashi, I. *et al.* *Nature* **512**, 319–323 (2014).
3. Bertrand, J. Y. *et al.* *Nature* **464**, 108–111 (2010).
4. Kissa, K. & Herbomel, P. *Nature* **464**, 112–115 (2010).
5. Lam, E. Y., Hall, C. J., Crosier, P. S., Crosier, K. E. & Flores, M. V. *Blood* **116**, 909–914 (2010).
6. Boisset, J. C. *et al.* *Nature* **464**, 116–120 (2010).
7. Pardanaud, L. *et al.* *Development* **122**, 1363–1371 (1996).
8. Pouget, C., Gautier, R., Teillet, M. A. & Jaffredo, T. *Development* **133**, 1013–1022 (2006).
9. Svetic, V. *et al.* *Development* **134**, 1011–1022 (2007).
10. Esner, M. *et al.* *Development* **133**, 737–749 (2006).
11. Mayeux-Louchart, A. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 8844–8849 (2014).
12. Clements, W. K. *et al.* *Nature* **474**, 220–224 (2011).
13. Sato, Y. *et al.* *Dev. Cell* **14**, 890–901 (2008).

This article was published online on 13 August 2014.

#### EARTH SCIENCE

## Warning signs of the Iquique earthquake

**An earthquake off Chile in 2014 occurred in a region where a great seismic event was expected. Two studies reveal that months of foreshocks and slow slip on the associated plate-boundary fault preceded the event. SEE LETTERS P.295 & P.299**

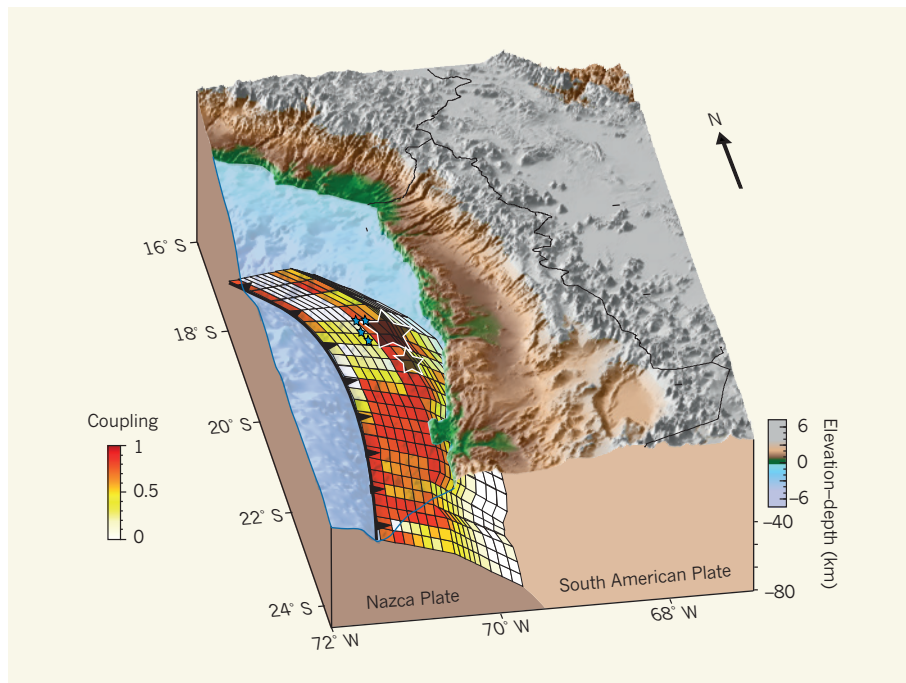
**ROLAND BÜRGMANN**

The most common question asked of an earthquake researcher is "When is the next Big One?" In short, the answer is "We don't know". But there is evidence that, at least in some cases, the answer could be more precise. In this issue, two studies<sup>1,2</sup> of the earthquake (magnitude 8.2) that occurred at Iquique, northern Chile, on 1 April 2014 suggest that a range of geophysical measurements collected in recent years indicated both a high overall earthquake probability and an increased short-term hazard in the region.

The Iquique event was a subduction earthquake — it occurred on the fault along which the oceanic Nazca Plate to the west thrusts itself below the South American continent at an average rate of about 7 centimetres per year. The last time a great earthquake occurred on this section of the plate boundary was in 1877, when a much larger event

(magnitude 8.6–8.8) ruptured nearly 500 kilometres of the subduction thrust fault. Measurements<sup>3,4</sup> of deformation at Earth's surface taken by the Global Positioning System (GPS) show that much of the fault that ruptured in 1877 is currently fully coupled (locked in position, and thus building up stress and slip deficit that will be released in a future earthquake; Fig. 1). This section of the plate boundary was therefore recognized as a seismic gap, a region of an active fault that seems to be overdue for one or more great earthquakes.

The Iquique earthquake occurred within this seismic gap, but was not nearly big enough to fill it. Hayes *et al.*<sup>1</sup> (page 295) and Schurr *et al.*<sup>2</sup> (page 299) constrained models of the earthquake slip using seismic data from local and global stations, together with geodetic measurements of surface deformation. They report that slip of up to about 5 m occurred in a zone stretching from the earthquake's focus in the north to the Chilean coast in the southeast.



**Figure 1 | Tectonic setting of the Iquique earthquake, 1 April 2014.** The diagram shows the Nazca Plate thrusting eastward below South America. The colours of the rectangular fault elements indicate the degree of coupling inferred from satellite-measured surface displacements in northern Chile. A coupling value of 1 (red) means that the fault is completely locked and builds up a slip deficit until the next earthquake. A low coupling value indicates that the plate-boundary fault slips aseismically. The black stars show the approximate extent of the mainshock (magnitude 8.2, large star) and its largest aftershock (magnitude 7.6, smaller star). Blue stars indicate the area of foreshock activity near the focus of the mainshock in the months before the rupture. Two papers<sup>1,2</sup> suggest that geophysical data collected in recent years indicated a high overall earthquake probability and an increased short-term hazard. (Figure modified from ref. 4.)

Two days later, an aftershock of magnitude 7.6 expanded the rupture zone to the south for a total length of about 200 km.

Of particular note was a period that lasted for at least three months, in which foreshocks propagated towards the eventual focus of the mainshock<sup>5</sup>. That is, rather than snapping with no warning signs, this great earthquake was preceded by a fascinating sequence of foreshocks that, in retrospect, can be understood as part of a slow unfastening process leading up to and triggering the eventual earthquake rupture. The foreshocks occurred in a zone that had previously been recognized<sup>3,4</sup> as being less strongly coupled — in which the fault slips slowly without causing an earthquake (Fig. 1). It seems that the foreshocks, accompanied by slow aseismic slip in this partially locked zone, ultimately initiated a dynamic earthquake rupture, breaking the fully locked section to the southeast.

There are still some questions to resolve regarding the sequence of events leading up to the Iquique earthquake. Schurr *et al.* find that the total surface displacements produced by models of the catalogued foreshocks in the second half of March 2014 match those observed with GPS. This indicates that there was little, if any, aseismic fault slip associated with this activity. By contrast, an independent analysis<sup>6</sup> of GPS data suggests that aseismic slow slip

in the foreshock region greatly exceeded slip associated with the foreshocks alone. Additional evidence for substantial aseismic fault creep comes from observations<sup>5</sup> of very small, identically repeating earthquakes among the foreshocks on the plate-boundary fault. Questions also remain about whether deformation within overlying crustal rocks contributed to the precursory activity, in addition to slip on the subduction thrust. The largest foreshock (magnitude 6.7) and several smaller events apparently occurred in the South American crust<sup>1,2</sup>, suggesting a complicated sequence of events leading up to the mainshock.

Most large earthquakes on plate-boundary faults are preceded by foreshock activity in the weeks before the event<sup>7</sup>. So should researchers have anticipated the Iquique event and provided some warning as foreshocks unfolded in early 2014? Such sequences do not yet allow for confident earthquake prediction, because there is no accepted or consistent pattern of activity before an impending large earthquake. Indeed, we still do not know how to recognize foreshocks as such when they occur. However, it seems that swarms of events that accompany transient slow slip near strongly locked sections of a fault, as apparently occurred before the Iquique earthquake, are more likely than most background earthquakes to be foreshocks of a large mainshock<sup>8</sup>.

Hayes *et al.* argue that if we can characterize the progression of both slow and earthquake slip on a plate boundary from high-quality geophysical data, we can also model the time-dependent rise in stress on the locked sections of the fault, and therefore formally estimate the increase in earthquake probability. For example, calculations have been made<sup>9</sup> of the changes of stress, and of the related increase in probability of a large earthquake, associated with a flurry of small earthquakes and associated slow slip for the locked section of the Hayward fault in California in 2011 and 2012. In this case, the short-term increase of seismic hazard from the section that last ruptured in 1868 was small. Such modelling of time-dependent deformation, stress and hazard might form the basis for time-dependent, operational earthquake forecasting<sup>10</sup>, and thus formalize the message embedded in such potentially precursory activity.

Little would have been known of the events leading up to the Iquique earthquakes if it had not been for the recent deployment of modern geodetic and seismic instrumentation in the region. Nonetheless, given that much of the activity occurred well offshore near the trench of the subduction zone, the distribution of land-based stations is sub-optimal. It is important to improve geodetic and seismic monitoring, and to include offshore sea-floor instrumentation<sup>11</sup>, so that we can better understand unfolding plate-boundary fault activity preceding some great earthquakes.

Comparison of the detailed models of the fault slip during the Iquique sequence<sup>1,2,6</sup> with the extent of the fully locked portions of the subduction thrust<sup>3,4</sup> worryingly indicates that only a small fraction of the seismic gap ruptured. As the current studies conclude, the Big One may still be to come. ■

**Roland Bürgmann** is in the Department of Earth & Planetary Science, University of California, Berkeley, Berkeley, California 94720-4767, USA.

e-mail: [burgmann@seismo.berkeley.edu](mailto:burgmann@seismo.berkeley.edu)

- Hayes, G. P. *et al.* *Nature* **512**, 295–298 (2014).
- Schurr, B. *et al.* *Nature* **512**, 299–302 (2014).
- Métis, M. *et al.* *Geophys. J. Int.* **194**, 1283–1294 (2013).
- Béjar-Pizarro, M. *et al.* *Nature Geosci.* **6**, 462–467 (2013).
- Kato, A. & Nakagawa, S. *Geophys. Res. Lett.* <http://dx.doi.org/10.1002/2014GL061138> (2014).
- Ruiz, S. *et al.* *Science* <http://dx.doi.org/10.1126/science.1256074> (2014).
- Bouchon, M., Durand, V., Marsan, D., Karabulut, H. & Schmittbuhl, J. *Nature Geosci.* **6**, 299–302 (2013).
- Brodsky, E. E. & Lay, T. *Science* **344**, 700–702 (2014).
- Shirzaei, M., Taira, T. & Bürgmann, R. *Earth Planet. Sci. Lett.* **371–372**, 59–66 (2013).
- Jordan, T. H. & Jones, L. M. *Seismol. Res. Lett.* **81**, 571–574 (2010).
- Bürgmann, R. & Chadwell, C. D. *Annu. Rev. Earth Planet. Sci.* **42**, 509–534 (2014).

This article was published online on 13 August 2014.



# The time of the last Neanderthals

The application of improved radiocarbon-dating techniques to samples from archaeological sites ranging from Russia to Spain has redefined the timing of the final disappearance of Neanderthals from Europe. [SEE LETTER P.306](#)

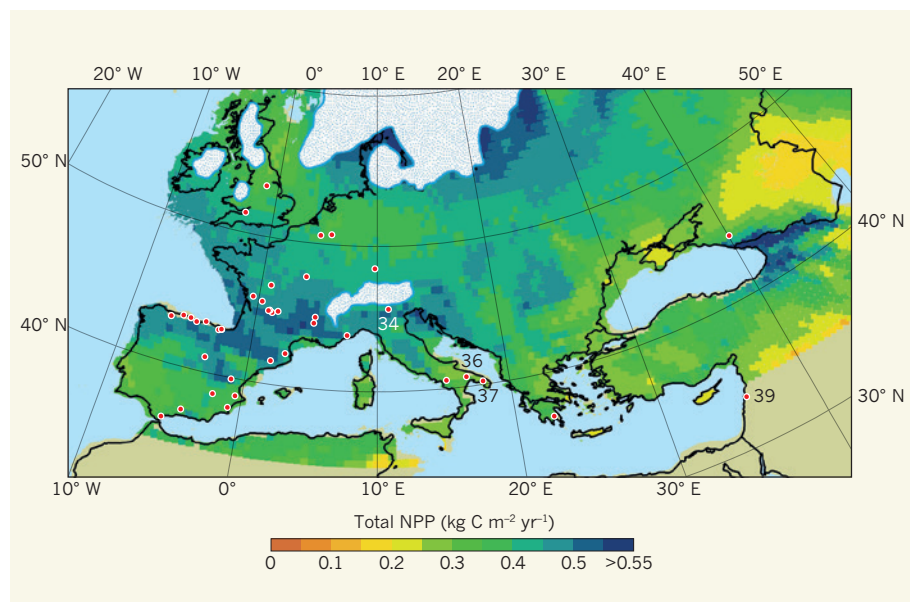
WILLIAM DAVIES

The question of when and how the Neanderthals became extinct has long fascinated scientists. Three key advances have begun to transform our understanding of this process: better removal of contaminants from ancient samples, more targeting of human fossils and artefacts for dating, and extension of calibration curves to 50,000 years ago for radiocarbon dates. On page 306 of this issue, Higham *et al.*<sup>1</sup> combine these developments to investigate how the biological and cultural transition from Neanderthals to modern humans might have occurred. The authors conclude that Neanderthals and their associated archaeological industries were gone from Europe between 41,000 and 39,000 years ago, leaving the continent to our species.

Improvements in methods for pretreating samples before radiocarbon dating, including the removal of humic-acid contaminants from charcoal<sup>2</sup> and the ultrafiltering of non-degraded collagen molecules from

bone samples<sup>3</sup>, have increased confidence in the accuracy and precision of radiocarbon dates older than 25,000 years ago. In cases in which ultrafiltration did not purify the sample satisfactorily, amino acids characteristic of collagen can be extracted for dating<sup>4</sup>. The resulting dates have given us a finer-grained picture of changes in human behaviours and species in the key period between 50,000 and 30,000 years ago, when modern humans (*Homo sapiens*) reached Europe and, eventually, replaced Neanderthals.

Higham and colleagues applied such pretreatments and dating methods to carefully selected samples from 40 sites. The samples included Neanderthal bones and artefacts from the Mousterian and Châtelperronian stone-tool industries, which are frequently associated with Neanderthals, and Uluzzian artefacts, possibly made by modern humans (Fig. 1). The authors then used Bayesian age models to convert their radiocarbon dates to calendar years, using stratigraphic positions of samples from each site to refine calibrated ages.



**Figure 1 | Ancient environments.** The map shows the location of the 40 sites from which Higham *et al.*<sup>1</sup> collected and dated archaeological samples. Many of these coincide with regions of high estimated levels of net primary (plant) productivity (NPP; measured in kilograms of carbon per square metre per year) for 42,000 years ago<sup>13</sup>. Most are thought to be Neanderthal sites; some (34, 36, 37, 39) may represent modern humans. White shading depicts ice cover.

According to the authors, Europe at 45,000 years ago was essentially Neanderthal, with small pockets of modern humans, represented by the Uluzzian stone-tool industry, in regions such as Italy. Their models plot changing spatial distributions of Neanderthals and modern humans over the subsequent five millennia, with temporal overlaps between Neanderthals and modern humans of between 470 and 4,900 years, depending on region. In human terms, these transitions spanned some 25 to 250 generations, culminating in the restriction of the Neanderthal presence in Europe to France at 40,000 years ago and their eventual extinction. Contrary to previous models<sup>5,6</sup>, the authors find no convincing evidence of Neanderthal survival in Iberia after 40,000 years ago.

The copious data collected by Higham *et al.* test traditional, assumed associations between human species and archaeological industries. The assumption that we can attribute Mousterian industries (which are based on triangular or roughly oval flakes derived from carefully shaped nodules) to Neanderthals is simplistic, because we know that modern humans made similar tools in the Levant and Africa<sup>7,8</sup>. Conversely, claims of a late Neanderthal presence in regions such as Gibraltar<sup>6</sup> have been based purely on the supposed exclusive association of European Mousterian industries with that species. There are no associated Neanderthal remains with such late (post-40,000 years ago) Mousterian industries, and thus the evidence for extended survival of Neanderthals in Iberia remains hypothetical.

Crude equivalences between human species and archaeological industries suited archaeologists when dating methods were similarly rudimentary. However, improvements in chronological accuracy and precision, and the incorporation of information derived from studies of ancient DNA and environments now allow more dynamic and mosaic models of change to be explored. 'Transitions' need not be seen as intermittent, sudden shifts from one set of stable conditions to another, but as dynamic processes operating at multiple spatial and temporal scales.

Higham *et al.* do not explore all the ramifications of their data. For example, morphological changes in human fossils can be directly dated and compared with time estimates for genetic evidence of interbreeding between Neanderthals and non-African modern humans. Currently, the greatest amount of gene flow is thought to have occurred around 77,000–114,000 years ago<sup>9</sup>, long before the claimed skeletal evidence for interbreeding at 45,000–35,000 years ago<sup>10</sup>. This seeming asynchronicity between genetic and skeletal patterning has yet to be explained.

In addition, the authors do not discuss why the Châtelperronian — a Neanderthal industry from France and northern Spain — apparently ended at much the same time as the

JUDY ALLEN

## MOLECULAR PHYSICS

modern-human Uluzzian industry of Italy and Greece, 39,000–40,000 years ago. Makers of the Uluzzian were perhaps partially affected by the nearby Campanian Ignimbrite volcanic eruption at around 39,400 years ago<sup>11</sup>, but we have no evidence of this eruption, and thus its associated cultural impact, farther west in France<sup>12</sup>. Thus, multiple reasons, including dispersal of new technologies (such as the Aurignacian) and modern-human groups, must be explored to explain the disappearance of these 'transitional' archaeological industries.

Spatial gaps also exist in Higham and colleagues' coverage: western Europe and the Mediterranean zones of eastern Europe are well represented, but inland central and eastern Europe are effectively excluded from their models. Archaeologically, the latter regions contain a diverse array of industries, often situated in patches of high plant productivity<sup>13</sup> (Fig. 1), the chronology and human-fossil associations of which are poorly understood. To assess further the behavioural, ecological and biological mosaics at the time of the transition from Neanderthals to modern humans, we need to bring this region up to the precision and coverage of the rest of Europe.

More chronological work is now required: new sites need to be dated, and reanalyses are required for problematic samples. Higham and colleagues' study has thrown down the gauntlet, and future researchers will need to try hard to demonstrate Neanderthal survival in Europe after 40,000 years ago. Researchers should instead focus on the more interesting questions of how and why human species and behaviours varied spatio-temporally in Europe in the period between 50,000 and 30,000 years ago, because understanding such patterns may provide fascinating insight into human ecology, social networks and the exchange of ideas at that time. ■

**William Davies** is in the Centre for the Archaeology of Human Origins, Faculty of Humanities, University of Southampton, Southampton SO17 1BF, UK.  
e-mail: s.w.g.davies@soton.ac.uk

1. Higham, T. *et al.* *Nature* **512**, 306–309 (2014).
2. Bird, M. I. *et al.* *Radiocarbon* **41**, 127–140 (1999).
3. Bronk Ramsey, C., Higham, T., Bowles, A. & Hedges, R. *Radiocarbon* **46**, 155–163 (2004).
4. Marom, A., McCullagh, J. S., Higham, T., Sinitsyn, A. A. & Hedges, R. E. *Proc. Natl Acad. Sci. USA* **109**, 6878–6881 (2012).
5. Zilhão, J. in *El Origen del Hombre Moderno en el Suroeste de Europa* (ed. Cabrera Valdés, V.) 127–145 (UNED, 1993).
6. Finlayson, C. *et al.* *Nature* **443**, 850–853 (2006).
7. McBrearty, S. & Brooks, A. J. *Hum. Evol.* **39**, 453–563 (2000).
8. Stringer, C. B., Grün, R., Schwarz, H. P. & Goldberg, P. *Nature* **338**, 756–758 (1989).
9. Prüfer, K. *et al.* *Nature* **505**, 43–49 (2014).
10. Trinkaus, E. *Proc. Natl Acad. Sci. USA* **104**, 7367–7372 (2007).
11. De Vivo, B. *et al.* *Miner. Petrol.* **73**, 47–65 (2001).
12. Lowe, J. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 13532–13537 (2012).
13. Allen, J. R. M. *et al.* *Quat. Sci. Rev.* **29**, 2604–2618 (2010).

# Complexity trapped by simplicity

**Devices known as magneto-optical traps have long been used to cool and confine atoms, but not molecules — until now. This new ability should enable many studies and applications of the physics of ultracold molecules. SEE LETTER P.286**

FRANCESCA FERLAINO

On page 286 of this issue, Barry *et al.*<sup>1</sup> report the first demonstration that diatomic molecules can be caught in a three-dimensional magneto-optical trap — a device that combines the effects of lasers and of a magnetic field to capture and cool particles. By using this approach, the authors reach the lowest temperature yet achieved for molecules using direct-cooling methods. This opens the way to the study of a plethora of fascinating phenomena in quantum physics, and might enable applications ranging from quantum information processing and simulations, to precision measurements and ultra-cold chemistry<sup>2</sup>.

Molecules — even simple diatomic molecules — are many times more complex than atoms. They possess a large number of internal states because of their electronic, vibrational and rotational degrees of freedom. Diatomic molecules can also carry an electric dipole moment and can undergo involved reactive processes, adding to their complexity. But it is precisely this complexity that makes molecules so interesting.

Tremendous progress has been made in cooling and slowing ensembles of molecules<sup>2</sup> to study their quantum behaviour. However, despite intense experimental and theoretical studies, developing a versatile strategy to simultaneously deeply cool and trap molecules of various kinds remains a major challenge. With the approach of the 20th anniversary of the first realization of an atomic Bose–Einstein condensate<sup>3</sup> — an ensemble of ultracold atoms that exhibits collective quantum behaviour — now is a good time to ask why molecules have been so much more difficult to capture and manipulate in cooling experiments.

The rapid development of ultracold-atom physics is certainly due in part to the success and robustness of laser cooling<sup>4,5</sup>. When it was first proved that massless photons could exert a radiation-pressure force on massive particles, such as sodium atoms, causing the atoms to slow down (that is, to be cooled), the full implications for atomic, molecular and optical physics were no doubt unseen. But in less than half a century, physicists have used this phenomenon in methods to manipulate, trap

and cool an ever-increasing number of atomic species, and have also trapped ions to ultracold temperatures. Nowadays, every standard experiment with ultracold atoms employs a magneto-optical trap (MOT) to capture and cool atoms down to temperatures of roughly tens of microkelvins.

The MOT technique was originally proposed by the physicist Jean Dalibard in the 1980s (ref. 6), and was demonstrated in the laboratory soon after<sup>7</sup>. A typical MOT is realized using three pairs of counter-propagating light beams and a static quadrupole magnetic field. Its working principle is based on laser cooling, which involves an exchange of momentum between an atom and a photon in repeated cycles of photon absorption and emission. During a cycle, the atom is first excited to a high-energy state and then spontaneously decays back to its initial state. Although the cooling effect is extremely powerful, it does not provide spatial confinement. But in a magnetic-field gradient, the light field becomes position dependent, and generates a restoring force that gathers atoms around the zero-magnetic-field position.

So how does this technique apply to the cooling and trapping of molecules, with their complicated internal states? The resulting complexity is so great that finding transitions between states suitable for absorption–emission cycles seems, at first glance, to be a hopeless task. This prejudice has been partially broken down by the experimental demonstration of direct laser cooling of some simple diatomic molecules<sup>8–10</sup>, including two-dimensional transversal MOT cooling of a molecular beam<sup>9</sup>, and cooling of polyatomic molecules through a mechanism known as the opto-electric Sisyphus effect<sup>11</sup>.

Barry and colleagues' report of magneto-optical trapping of strontium monofluoride molecules represents a major advance in molecular cooling. Using their approach, complex molecules can be treated like atoms — in other words, the standard three-dimensional configuration of an atomic MOT can be applied to molecules.

The researchers' optical-cooling cycle uses a remarkably simple recycling scheme that involves just a few excited molecular states<sup>12</sup> and only four light wavelengths. The trick is



to use a manifold of rotational excited states that has fewer energy levels than the ground-state manifold. The molecules therefore have relatively few escape routes from the cooling cycle, so that, in this case, only three light wavelengths are needed to pump escaped atoms back into the cycle. The price paid for this approach is a weaker net radiation-pressure force than that in ordinary atomic MOTs, and thus fewer molecules collected in the trap. However, schemes and ideas of how to circumvent this problem have been proposed<sup>9</sup>.

Barry *et al.* report that, after  $10^6$  cycles of absorption and emission, their MOT contains about 300 molecules at a temperature of 2.5 millikelvins. These numbers, although certainly below the performance of ordinary MOTs of atoms (which typically trap  $10^6$  to  $10^9$  atoms at temperatures of tens to hundreds of microkelvins) prove for the first time that molecules can be cooled and trapped. Cooling decreases the range of velocities of molecules

in an ensemble, whereas trapping clusters the molecules together. This combination provides a route to high-density molecular ensembles.

The magneto-optical trapping of molecules might have the same tremendous impact as its atomic counterpart, revolutionizing the field of molecular cooling. Barry and co-workers' strategy of applying atomic approaches to molecules is only the beginning. As noted earlier, molecules possess a feature absent in atoms: strong electric dipole moments. It is time to use this property at the lowest temperatures achievable in MOTs to trigger dipolar scattering, or to create a reservoir of electrically trapped molecules in metastable states to refill the MOT, as has been achieved for strongly magnetic atoms by exploiting their magnetic dipoles<sup>13</sup>. ■

**Francesca Ferlaino** is at the Institute for Experimental Physics, University of Innsbruck, Austrian Academy of Sciences, Innsbruck 6020, Austria.

e-mail: francesca.ferlaino@uibk.ac.at

1. Barry, J. F., McCarron, D. J., Norrgard, E. B., Steinecker, M. H. & DeMille, D. *Nature* **512**, 286–289 (2014).
2. Carr, L., DeMille, D., Kreams, R. & Ye, J. *New J. Phys.* **11**, 055049 (2009).
3. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).
4. Hänsch, T. W. & Schawlow, A. *Opt. Commun.* **13**, 68–69 (1975).
5. Cohen-Tannoudji, C. & Guéry-Odelin, D. *Advances in Atomic Physics: An Overview* (World Scientific, 2011).
6. www.nobelprize.org/nobel\_prizes/physics/laureates/1997/chu-lecture.pdf
7. Raab, E. L., Prentiss, M., Cable, A., Chu, S. & Pritchard, D. E. *Phys. Rev. Lett.* **59**, 2631 (1987).
8. Shuman, E. S., Barry, J. F. & DeMille, D. *Nature* **467**, 820–823 (2010).
9. Hummon, M. T. *et al. Phys. Rev. Lett.* **110**, 143001 (2013).
10. Zhelyazkova, V. *Phys. Rev. A* **89**, 053416 (2014).
11. Zeppenfeld, M. *et al. Nature* **491**, 570–573 (2012).
12. Stuhl, B. K., Sawyer, B. C., Wang, D. & Ye, J. *Phys. Rev. Lett.* **101**, 243002 (2008).
13. McClelland, J. & Hanssen, J. *Phys. Rev. Lett.* **96**, 143005 (2006).

might conclude that Austronesian languages arose in Indonesia. But that would be like mistakenly concluding that modern Germanic languages (the language subfamily that includes English) arose in North America, where most Germanic-language speakers live today. Instead, linguists deduce that Austronesian languages arose on Taiwan, because eight of the approximately nine Austronesian-language subfamilies are confined to Taiwan's aboriginal people (its inhabitants before the Chinese immigration). All Austronesian languages outside Taiwan belong to the ninth subfamily.

This picture, of a spread of Austronesian languages out of Taiwan, resembles the known history of Germanic languages, which arose in continental northwest Europe and still have their highest diversity there, and of which only one spread to England and thence to North America. Could the postulated spread of Austronesian languages from Taiwan through Indonesia be related to the spread of an ancient Chinese people through Indonesia, as suggested by human biology and ancient skeletons?

Further evidence comes from archaeology<sup>5,6</sup>. This tells us that, until about 2200 BC, all known Philippine and Indonesian populations were hunter-gatherers without pottery, farming or polished stone tools. Beginning around 2200 BC, red pottery, Taiwan jade, polished stone tools and domestic animals spread across the Luzon Strait from Taiwan, first to the northern Philippines and then south and east through Indonesia. Might those first Philippine and Indonesian farmers have carried not only animals and human skeletal traits but also Austronesian languages?

Enter Lipson and colleagues, whose genetic evidence<sup>4</sup> clarifies previous genetic studies<sup>7–10</sup>.

## POPULATION HISTORY

# Human melting pots in southeast Asia

**New genetic methods to analyse mixed human populations have extended existing, multidisciplinary evidence for the historical migrations and mixings of Austronesian peoples.**

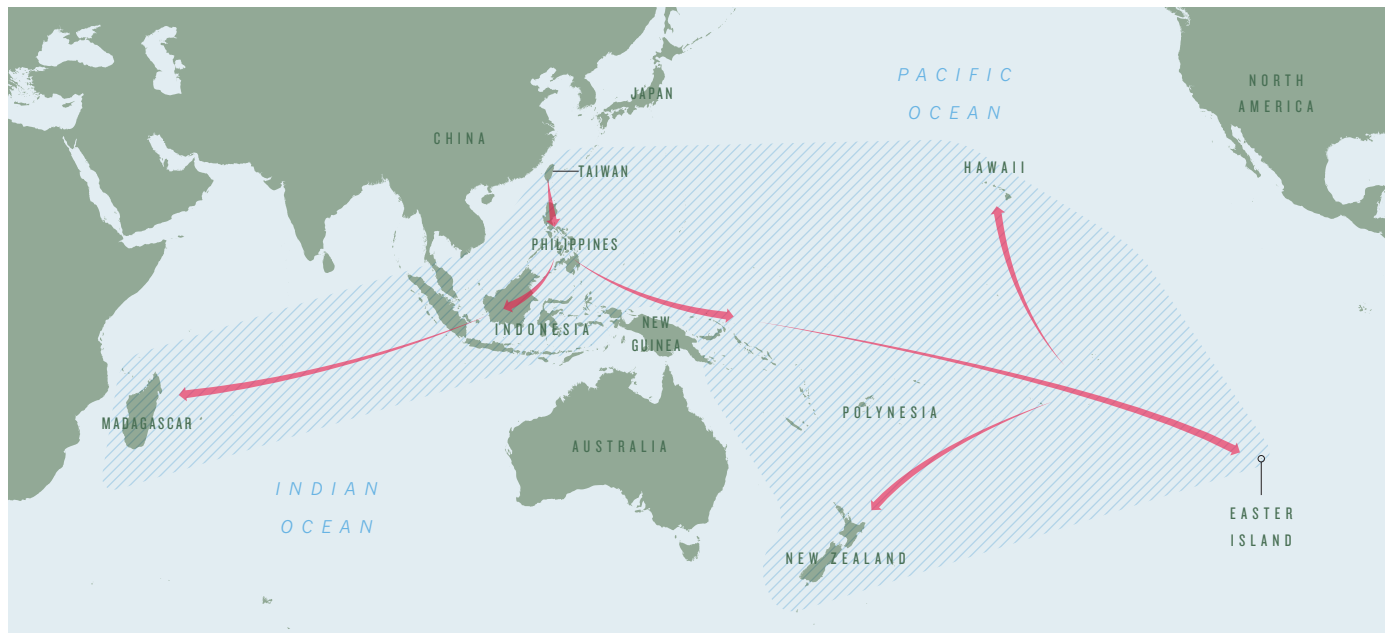
JARED DIAMOND

Five centuries ago, the world's most widespread language family was the Austronesian family, which includes modern Philippine, Indonesian and Polynesian languages<sup>1–3</sup>. It extended halfway around the globe, from Madagascar to Easter Island (Fig. 1). How did these languages spread so far, before colonial empires and modern ships existed? Writing in *Nature Communications*, Lipson *et al.*<sup>4</sup> present comprehensive genetic data that, combined with previous studies<sup>5,6</sup> of human biology, ancient skeletons, linguistics and archaeology, build a picture of Austronesian migrations and ancestry. In the process, they introduce methods for resolving components of genetically mixed human populations, which constitute most populations today. As a result, the Austronesian story becomes a model of a multidisciplinary approach to reconstructing human history.

Modern populations in mainland and island southeast Asia fall into three groups as classified by differences in human biology — physical characteristics and appearance. The largest group encompasses modern Chinese

and most mainland southeast Asians, Philippine people and Indonesians west of New Guinea. A second group, the Australo-Papuans, consists of modern Papuans (also known as New Guineans) and the distantly related Aboriginal Australians. The third group (Negritos) comprises diminutive Papuan-like peoples scattered from the Andaman Islands to the Philippines. From ancient skeletons, we know that, until 5,000 years ago, the people of south China, mainland southeast Asia and Indonesia were Australo-Papuans or Negritos, rather than Chinese-like. Hence, these first two strands of evidence — human biology and ancient skeletons — suggest that people from China replaced the original populations of mainland southeast Asia and Indonesia in the past 5,000 years.

The linguistic evidence is dramatic<sup>1–3</sup>. All modern Philippine, Indonesian and Polynesian languages, some languages of coastal New Guinea and southeast-most mainland Asia, and Madagascar's Malagasy language are Austronesian. From these facts, a non-linguist, inclined to guess that a language family would have originated near the centre of its modern geographic range and populations,



**Figure 1 | The modern geographic extent (shaded) of Austronesian languages, and their proposed routes of spread.**

The authors used single-nucleotide variations in DNA sequences to construct a phylogenetic tree of 31 Austronesian and 25 non-Austronesian populations. They then developed a quantitative method, termed MixMapper 2.0, to identify and statistically test multiple source populations contributing to this tree.

Their analysis of Austronesian populations reveals four ancestral source components, most-closely related to aboriginal Taiwanese people, Negritos, Papuans and H'tin (a mainland southeast-Asian people speaking an Austro-Asiatic language). The aboriginal Taiwanese seem to consist of only the first component, but all other Austronesian populations analysed were found to be mixtures of two or three components: always the Taiwanese-like component, plus one or two others. The second-largest component is the Negrito-like in the Philippines, the Papuan in east Indonesia, and Polynesia and the Austro-Asiatic in west Indonesia.

The first three of these four components support indications from human biology, skeletons and archaeology that ancient south Chinese farmers colonized Taiwan and overran previous hunter-gatherer populations of the Philippines and Indonesia, in the process incorporating their genes (especially genes from Negritos in the north and west and from Papuans in the east). But the biggest surprise of Lipson and colleagues' study is the identification of a genetic component (especially in west Indonesia) related to speakers of Austro-Asiatic languages. Today, these languages are confined almost exclusively to mainland southeast Asia; except for modern Vietnamese and Cambodian, they are mainly spoken in small pockets surrounded by speakers of Tai and Sino-Tibetan languages. Evidently, the first farmers who spread south from China's

agricultural homeland spoke Austro-Asiatic languages, which then became partly replaced by other languages carried by subsequent farmer expansions.

How did genes of mainland Asian farmers speaking Austro-Asiatic languages end up in west Indonesia, where no Austro-Asiatic languages are spoken today? The most probable scenario is that Taiwan's Austronesian farmers spread not only south into the Philippines, but also southwest into mainland southeast Asia, mixed there with Austro-Asiatic farmers and then colonized west Indonesia. But Lipson *et al.* acknowledge two other possibilities: that Austronesians spreading south encountered Austro-Asiatic farmers already established in west Indonesia, for whom archaeological evidence is still lacking; or that Austro-Asiatic farmers spread into west Indonesia after Austronesian farmers arrived there.

Lipson and colleagues' conclusions focus attention on fascinating unresolved questions. For example, which of these hypotheses for how Austronesians picked up Austro-Asiatic genes is correct? How and when did Chinese languages replace proto-Austronesian languages in coastal South China? Why do all Philippine and Indonesian populations today, which are genetic mixtures of Austronesians and non-Austronesians, speak Austronesian languages, except for a few Papuan languages in Timor and Halmahera? And do modern Austronesians evince any genetic contribution from the micropigmies<sup>11</sup> who formerly inhabited the island of Flores and became replaced there by *Homo sapiens* hunter-gatherers, who in turn mixed with Austronesian farmers?

Two further questions involve specific peoples of perplexing origins. The most astonishing fact of modern human geography is

Madagascar's colonization around AD 500, from across the Indian Ocean, by Austronesian farmers speaking a language whose closest relative today is spoken on the Indonesian island of Borneo. But the Malagasy's ancestors surely did not arrive by a non-stop 8,000-kilometre flight from Borneo; will future genetic studies illuminate the intermediate steps? Finally, Austronesians expanded from Taiwan south into the Philippines, east to the Marianas and southwest to Malaysia. Might they also have expanded north to Japan, thereby contributing to the much-debated origins of the Japanese people<sup>12</sup>? ■

**Jared Diamond** is in the Department of Geography, University of California, Los Angeles, Los Angeles, California 90095-1524, USA.  
e-mail: [jdiamond@geog.ucla.edu](mailto:jdiamond@geog.ucla.edu)

1. Blust, R. *The Austronesian Languages* (Pacific Linguistics, 2009).
2. Pawley, A. in *Examining the Farming/Language Dispersal Hypothesis* (eds Bellwood, P. & Renfrew, C.) 251–274 (McDonald Inst., 2002).
3. Gray, R. D., Drummond, A. J. & Greenhill, S. J. *Science* **323**, 479–483 (2009).
4. Lipson, M. *et al. Nature Commun.* **5**, 4689; <http://dx.doi.org/10.1038/ncomms5689> (2014).
5. Bellwood, P. *First Migrants: Ancient Migration in Global Perspective* (Wiley-Blackwell, 2013).
6. Bellwood, P. *First Farmers: The Origins of Agricultural Societies* (Wiley-Blackwell, 2004).
7. Oppenheimer, S. J. & Richards, M. *Nature* **410**, 166–167 (2001).
8. Jinam, T. A. *et al. Mol. Biol. Evol.* **29**, 3513–3527 (2012).
9. Friedlaender, J. S. *et al. PLoS Genet.* **4**, e19 (2008).
10. Xu, S. *et al. Proc. Natl Acad. Sci. USA* **109**, 4574–4579 (2012).
11. Morwood, M. & van Oosterzee, P. *The Discovery of the Hobbit: The Scientific Breakthrough that Changed the Face of Human History* (Random House Australia, 2007).
12. Kumar, A. *Globalizing the Prehistory of Japan: Language, Genes and Civilization* (Routledge, 2009).



# Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway

Ashton Trey Belew<sup>1\*</sup>, Arturas Meskauskas<sup>1,2\*</sup>, Sharmishtha Musalgaonkar<sup>1</sup>, Vivek M. Advani<sup>1</sup>, Sergey O. Sulima<sup>1†</sup>, Wojciech K. Kasprzak<sup>3</sup>, Bruce A. Shapiro<sup>4</sup> & Jonathan D. Dinman<sup>1</sup>

**Programmed –1 ribosomal frameshift (–1 PRF) signals redirect translating ribosomes to slip back one base on messenger RNAs. Although well characterized in viruses, how these elements may regulate cellular gene expression is not understood. Here we describe a –1 PRF signal in the human mRNA encoding CCR5, the HIV-1 co-receptor. CCR5 mRNA-mediated –1 PRF is directed by an mRNA pseudoknot, and is stimulated by at least two microRNAs. Mapping the mRNA–miRNA interaction suggests that formation of a triplex RNA structure stimulates –1 PRF. A –1 PRF event on the CCR5 mRNA directs translating ribosomes to a premature termination codon, destabilizing it through the nonsense-mediated mRNA decay pathway. At least one additional mRNA decay pathway is also involved. Functional –1 PRF signals that seem to be regulated by miRNAs are also demonstrated in mRNAs encoding six other cytokine receptors, suggesting a novel mode through which immune responses may be fine-tuned in mammalian cells.**

Viral programmed ribosomal frameshift events typically produce carboxy-terminally extended fusion proteins. However, computational analyses predict that >95% of –1 PRF events on cellular mRNAs direct ribosomes to premature termination codons (PTC), suggesting that –1 PRF may be used by cells to regulate gene expression by destabilizing mRNAs through the nonsense-mediated mRNA decay (NMD) pathway<sup>1</sup>. Whereas a role for –1 PRF has been shown in yeast<sup>2,3</sup>, it has not been tested in higher eukaryotes so far. In yeast, mutants and drugs that globally affect –1 PRF generally promote deleterious phenotypes<sup>4</sup>, and global dysregulation of –1 PRF may contribute to human disease<sup>3,5–7</sup>. How sequence-specific regulation of –1 PRF might be achieved has been the central unanswered question in the field.

## A –1 PRF signal in the CCR5 receptor mRNA

CCR5 is a cytokine receptor which is exploited by HIV-1 as a co-receptor for entry into CD4<sup>+</sup> T-cells<sup>8</sup>. A strong candidate –1 PRF signal beginning at nucleotide 407 in the human CCR5 mRNA was identified computationally<sup>9</sup>. This sequence is >99% conserved among the great apes and is highly conserved among the higher primates (Extended Data Fig. 1). Using dual luciferase reporters (Extended Data Fig. 2a), the CCR5 sequence promoted 9–11% –1 PRF in HeLa (Fig. 1a) and 4.5–6.3% in Chinese hamster ovary (CHO) or Vero cells (Extended Data Fig. 2b). Mutagenesis of the slippery site from UUUAAAA to GCGCGCG reduced –1 PRF to <1% (Fig. 1a). Introduction of an in-frame termination codon (PTC control) 5' of the CCR5-derived sequence, or placing the firefly luciferase reporter out of frame with respect to *Renilla* reduced –1 PRF levels by more than two orders of magnitude (Fig. 1a), ruling out the possible presence of either a splicing donor site or an internal ribosome entry signal (IRES). An *in vitro* translation assay revealed a peptide consistent with a CCR5 –1 PRF event at levels comparable to that promoted by the HIV-1 –1 PRF signal (Fig. 1b). Liquid chromatography dual mass spectroscopic analysis of an affinity

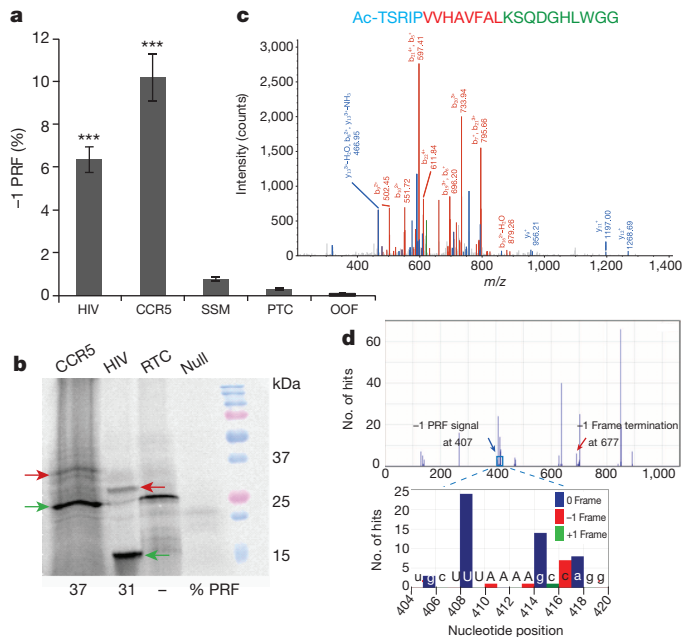
purified CCR5–β-gal fusion protein (Extended Data Fig. 2c) unambiguously identified the predicted –1 frameshift peptide harbouring the junction between the 0– and –1 frame encoded CCR5 sequence (Fig. 1c, Extended Data Fig. 2d). Analysis of published ribosome profiling data from human cells<sup>10</sup> revealed a sizable fraction of ribosomes paused at the CCR5 –1 PRF signal, 9/59 (~15%) of which were shifted into the –1 reading frame (Fig. 1d). These four lines of inquiry demonstrate that this sequence in the human CCR5 mRNA promotes efficient –1 PRF.

## Structural analysis

Computational analyses predicted the presence of two nearly equivalent downstream mRNA pseudoknots or a tandem stem-loop structure immediately 3' of the slippery site (Extended Data Fig. 3a). Analyses of chemical modification assays of a CCR5 runoff transcript (Extended Data Fig. 3b, c) were consistent with the presence of a two-stemmed mRNA pseudoknot (Fig. 2). Whereas the slippery site distal region of Stem 1 is stable, the proximal region is conformationally dynamic, consistent with single-molecule optical trap experiments revealing a complex network of folding pathways for this element<sup>11</sup>. The weak slippery site proximal half of stem 1 coupled with the internal bulge is reminiscent of the HIV-1 –1 PRF signal solution structure<sup>12</sup> and is consistent with the emerging view of conformational complexity as a critical feature of recoding pseudoknots<sup>13,14</sup>. Stem 2 contains four semi-helical segments (labelled a, b, d, e in Fig. 2a), plus a small segment in the middle (c), all separated by unpaired bases. The unpaired bases may allow the entire structure to bend, enabling U23 to bridge the gap between C22 and U24. The 'best fit' conformer diagrammed in Fig. 2a was used as the basis for molecular-dynamics-based simulation of the CCR5 –1 PRF stimulatory mRNA structure (Fig. 2b, Extended Data Fig. 3d, e). The root mean squared deviation (r.m.s.d.) average structure was calculated for the last 12 ns of an 80-ns long molecular dynamics simulation, where

<sup>1</sup>Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742, USA. <sup>2</sup>Department of Biotechnology and Microbiology, Vilnius University, Vilnius, LT 03101, Lithuania. <sup>3</sup>Basic Science Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, USA. <sup>4</sup>Basic Research Laboratory, National Cancer Institute, Frederick, Maryland 21702, USA. <sup>†</sup>Present address: VIB Center for the Biology of Disease, KU Leuven, Campus Gasthuisberg, Herestraat 49, bus 602, 3000 Leuven, Belgium.

\*These authors contributed equally to this work.



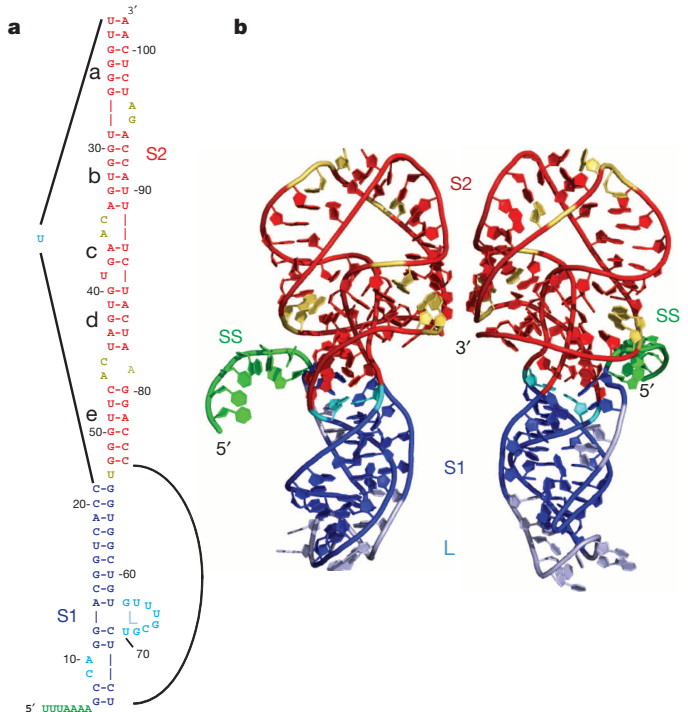
**Figure 1 | The CCR5 sequence promotes efficient frameshifting.**

**a**, Measurement of -1 PRF in HeLa cells. -1 PRF efficiency was monitored in HeLa cells using dual luciferase reporters. Error bars denote an approximation of standard errors. \*\*\* $P < 0.001$  compared to the out of frame control (Student's two-tailed  $t$ -test). **b**, Efficient -1 PRF promoted by the CCR5 sequence *in vitro*. Autoradiogram of *in vitro* translation reaction using mRNAs harbouring CCR5- or HIV-1-derived -1 PRF signals. Green arrows denote 0-frame encoded products. Red arrows denote -1 PRF encoded peptides. RTC indicates the readthrough control. Percentage -1 PRF promoted by CCR5 and HIV-1 frameshift signals is indicated below the lanes. **c**, Liquid chromatography with tandem mass spectrometry (LC-MS/MS) spectrum of a proteolytic fragment containing the CCR5 frameshift peptide. N-terminally acetylated leader peptide sequence is coloured blue, CCR5-derived 0-frame sequence beginning at V94 is red, and CCR5 -1 frame encoded sequence beginning after L101 is coloured green. **d**, Ribosomes accumulate at the CCR5 -1 PRF signal. Data mined from ref. 10. Top, locations of the -1 PRF signal and first -1 frame termination codon are indicated. Bottom, profiling data at the slippery site (indicated in capital letters) at single nucleotide resolution. Ribosomes arrested in the three different reading frames are colour-coded.

the r.m.s.d.s of the full structure and its two sub-domains SL1 (nucleotides 8–22 and 55–75) and SL2 (nucleotides 24–53 and 76–103) are most stable. The total energy for this structure is  $-24,296 \text{ kcal mol}^{-1}$ .

### Stimulation of -1 PRF by miR-1224

As cellular gene expression tends to be regulated, it is reasonable to hypothesize that -1 PRF might be regulated in a sequence-specific manner. This could be achieved through base-pairing interactions between specific small noncoding RNAs (ncRNAs) and -1 PRF signals, a hypothesis supported by the ability of antisense oligonucleotides to stimulate -1 PRF (reviewed in ref. 1). Computational searches revealed miR-1224, miR-711 and miR-141 as potential interacting partners with the CCR5 -1 PRF signal (Extended Data Fig. 4a). Transfection of HeLa cells with a miR-1224 precursor revealed concentration-dependent enhancement of CCR5-mediated -1 PRF (Fig. 3a). miR-1224 did not affect HIV-1-mediated -1 PRF. Addition of a miR-1224 antagonist (anti-miR-1224), or short interfering RNA knockdown of argonaute 1 reversed the effect of miR-1224 on CCR5-mediated -1 PRF (Fig. 3b). Although anti-miR-1224 seemed to stimulate CCR5 -1 PRF, the effect was not significant ( $P = 0.15$ ). siRNA knockdown of mRNAs encoding proteins involved in miRNA processing inhibited CCR5-mediated -1 PRF, but stimulated HIV-1-driven -1 PRF (Fig. 3c), supporting the model of sequence-specific regulation of -1 PRF through interactions



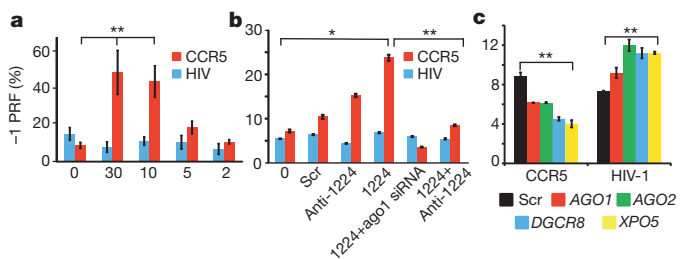
**Figure 2 | Models of the CCR5 -1 PRF stimulating mRNA pseudoknot.**

**a**, Best-fit two-dimensional model based on chemical modification analyses. **b**, Three-dimensional model, two views. Slippery site is green, stem 1 is dark blue, unpaired bases and the loop within stem 1 are light blue, paired bases in stem 2 are red, and unpaired bases in stem 2 are yellow.

between miRNAs and -1 PRF signals. Neither miR-141 nor miR-711 affected -1 PRF in a HeLa-cells-based assay (Extended Data Fig. 4b), perhaps owing to the presence of endogenous miR-711, and/or miR-141. However, miR-141 specifically stimulated CCR5-mediated -1 PRF in CHO cells (Extended Data Fig. 4c).

### miR-1224/CCR5 mRNA interactions

Two different *in vitro* electrophoretic mobility shift assays (EMSAs) were used to probe the interactions between the CCR5 -1 PRF signal and miR-1224. In one, the RNAs were mixed and incubated at physiological temperature ('native'), whereas in the second, they were co-denatured at high temperature and slowly annealed ('refolded'). Both



**Figure 3 | Specific stimulation of CCR5-mediated -1 PRF by miR-1224.**

**a**, HeLa cells were transfected with 0–30 nmol of miR-1224 miRNA expressing constructs and with HIV-1 or CCR5 -1 PRF reporters. **b**, PRF assays of HeLa cells mock-transfected (0), or transfected with scrambled miRNA (Scr), a miR-1224 antagonist (anti-1224), miR-1224, miR1224 + anti-miR-1224, or miR-1224 plus a siRNA directed against argonaute 1. **c**, Ablation of the miRNA processing machinery affects -1 PRF promoted by the HIV-1 and CCR5 frameshift signals. -1 PRF assays were performed using cells transfected with siRNAs targeting Argonaute 1 (AGO1), Argonaute 2 (AGO2), DGCR8, exportin 5 (XPO5) or scrambled sequences (Scr). Error bars denote standard error. \* $P < 0.05$ , \*\* $P < 0.01$  (Student's two-tailed  $t$ -test).



reactions were resolved through native polyacrylamide gel electrophoresis (PAGE; Extended Data Fig. 5a, b). Although miR-1224 interacted with the CCR5 sequence with sub-nanomolar dissociation constants in both conditions, its affinity was approximately twofold higher in the 'native' context (Fig. 4a). miR-1224 enhanced the appearance of multiple pre-existing conformers, particularly in the 'refolded' context, consistent with the structurally complex nature of the pseudoknot. miR-1224 did not interact with a transcript containing the HIV-1 -1 PRF signal (Extended Data Fig. 5c, d).

An affinity capture assay to probe CCR5-miR-1224 interactions in HeLa Tzm-BL cells expressing CCR5<sup>15</sup> revealed an approximately threefold enrichment for CCR5 mRNA relative to cells transfected with a scrambled control (Fig. 4b). In a parallel experiment in HeLa cells, the CCR5 -1 PRF signal containing dual-luciferase reporter mRNA was enriched more than 2,000-fold compared to no-miRNA controls (Fig. 4c), whereas the HIV-1 -1 PRF reporter was only enriched about tenfold. These findings demonstrate that miR-1224 specifically interacts with the CCR5 -1 PRF signal in live cells. Selective 2'-hydroxyl

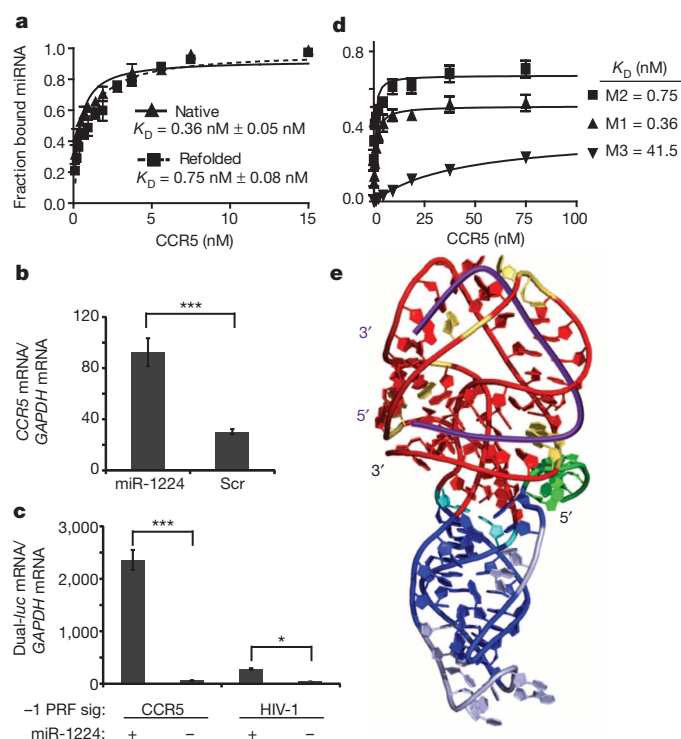
acylation analysed by primer extension (SHAPE) did not reveal differences in RNA modification patterns in the presence of miR-1224 (Extended Data Fig. 5e), suggesting that miR-1224 does not function to create any new conformation(s) of the CCR5 -1 PRF signal *per se*. Rather, it may stabilize a pre-existing structure(s) promoting efficient -1 PRF.

## Mapping the miR-1224 binding site

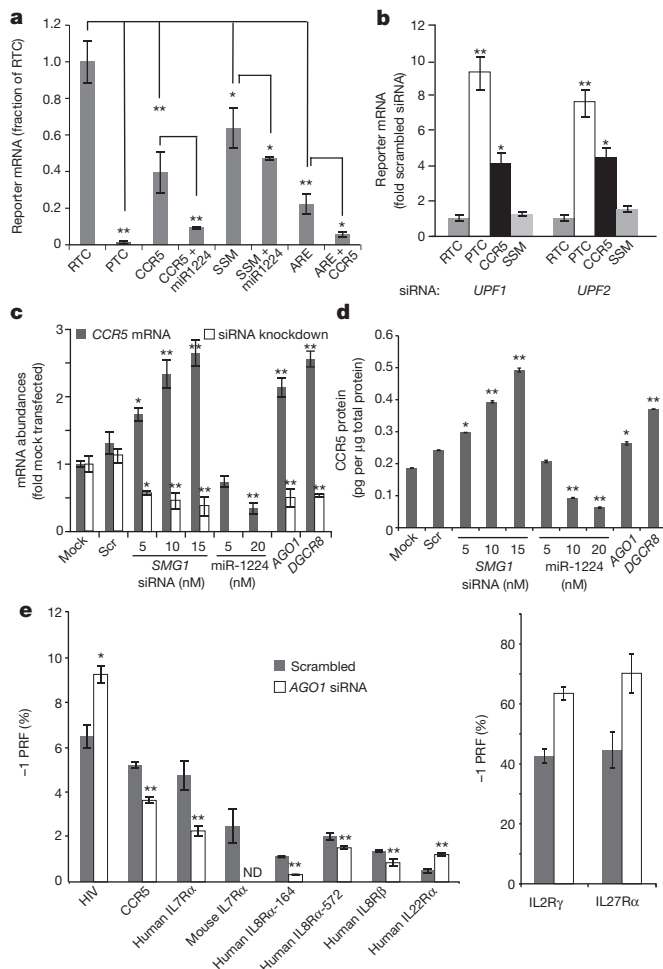
CCR5-derived transcripts harbouring mutations in the predicted miR-1224 interacting sequences (mutants M1-M3, Extended Data Fig. 6a) were assayed by EMSA. The 5' proximal mutant (M1) yielded the same  $K_D$  (0.76 nM) as the wild-type sequence under 'refolded' conditions, the central sequence mutant (M2) promoted the same dissociation constant as the wild-type sequence assayed under 'native' conditions (0.36 nM), and the 3' proximal binding site mutant (M3) caused an approximately 100-fold increase in  $K_D$  (42 nM) (Fig. 4d, Extended Data Fig. 6b, c). These findings suggest that miR-1224 may participate in a triple helical interaction with subdomains a-d of Stem 2 under native conditions (modelled in Fig. 4e and Extended Data Fig. 6c). The predicted triple-base interaction between miR-1224 and Stem 2 is consistent with the stable 3' end of the pseudoknot identified in the molecular dynamics simulation (Extended Data Fig. 3d, e). The 'torsional restraint' model of -1 PRF posits that ribosomes are directed to pause over the slippery site by Stem 2-induced supercoiling of Stem 1<sup>16</sup>. The miR-1224 mapping data are consistent with this model: increased stability of the Stem 2 by the mRNA-miRNA interaction renders this structure even more difficult to resolve, further increasing the fraction of paused ribosomes. miR-141 is predicted to interact with the same region of the CCR5 -1 PRF signal whereas miR-711 is not, suggesting that miR-141 enhances CCR5-directed -1 PRF in a similar manner to miR-1224. The sequence of the mature miR-1224 is 100% conserved among higher primates (*Homo*, *Pan*, *Pongo* and *Macaca*) as is its binding site with the 3' end of their respective CCR5 -1 PRF signals, suggesting that miR-1224-mediated regulation of CCR5 -1 PRF is evolutionarily conserved. It is also notable that the miR-1224/CCR5-interacting sequences do not conform to established seed sequences for miRNAs.

## mRNA suicide through -1 PRF

Ribosome profiling data also revealed a cluster of ribosomes paused at the first -1 frame termination codon after the CCR5 slippery site (Fig. 1d). A series of rabbit  $\beta$ -globin-derived reporters (Extended Data Fig. 7a) were used to assess the effects of the CCR5 -1 PRF signal on mRNA steady-state abundance and stability. Steady-state abundance of the CCR5 -1 PRF-containing reporter mRNA was about 38% of the readthrough control and was further decreased upon addition of miR-1224 (~10% of readthrough control), consistent with an inverse correlation between -1 PRF efficiency and mRNA abundance<sup>2</sup> (Fig. 5a). An in-frame PTC strongly decreased mRNA abundance (~1% of readthrough control). A reporter with the tumour necrosis factor (TNF)- $\alpha$ -derived AU-rich element (ARE) in its 3' untranslated region (UTR)<sup>17</sup> reduced mRNA abundance to ~22%. In combination with the CCR5 -1 PRF signal, mRNA abundance was reduced to ~6%, consistent with NMD and ARE-mediated decay operating independently of one another. The CCR5 slippery site mutant (SSM) decreased reporter mRNA abundance to ~64%, and addition of miR-1224 decreased this to ~47%. The former finding suggests that the stable mRNA pseudoknot has mRNA destabilizing activity independent of frameshifting, perhaps through the no-go mRNA pathway as described in yeast<sup>18</sup>. Its stabilization by miR-1224 may enhance this process. Alternatively, miR-1224 may promote accelerated mRNA turnover through canonical miRNA-mediated translational repression<sup>19</sup>. However, if this were true, miR-1224 should have reduced SSM mRNA abundance to the same extent as the native sequence. Abundance of the CCR5 -1 PRF signal containing reporter mRNA was increased by about 4.4-fold by partial siRNA knockdown of NMD (Fig. 5b). Abundance of the SSM construct was not increased by NMD ablation, consistent with its mRNA



**Figure 4 | Mapping and modelling the interactions of miR-1224 with the CCR5 -1 PRF signal *in vitro* and in live cells.** **a**, Dilutions of CCR5 -1 PRF signal (R5) transcript were mixed with [<sup>32</sup>P]-labelled miR-1224 RNA (miR), and incubated at 30 °C (Native), or denatured at 90 °C and slowly cooled (Refolded). Samples separated through native PAGE were quantified and data plotted onto single site binding isotherms.  $K_D$  values and standard deviations are indicated. **b**, *In vivo* pull-down of native CCR5 mRNA in live cells. Biotinylated miR-1224 precursor or a scrambled biotinylated control (Scr) were transfected into HeLa Tzm BL cells expressing CCR5. Fold enrichment of affinity purified mRNAs were analysed by quantitative PCR with reverse transcription (qRT-PCR) using CCR5- or GAPDH-specific primer sets. **c**, HeLa cells were co-transfected with dual-luciferase plasmids containing either the CCR5 or HIV-1 -1 PRF signal sequences and affinity-purified mRNAs were analysed as in **b**. **d**, EMSA assays were performed using miR1224 and M1, M2 and M3 variants of the CCR5 signal using native conditions. Single site binding isotherms generated from these data are plotted.  $K_D$  values are indicated. For **a** and **d**,  $n = 6$  for each sample (three times each of two technical replicates). For **b** and **c**,  $n = 9$  for each sample (three times each for three biological replicates). Error bars denote standard deviation. \* $P < 0.05$ , \*\*\* $P < 0.001$  (Student's two-tailed  $t$ -test). **e**, Conceptual model of CCR5 pseudoknot complexed with miR-1224 (purple).



**Figure 5 | miRNA mediated control of CCR5 expression through -1 PRF and NMD.** **a**, qRT-PCR analysis of rabbit  $\beta$ -globin reporter (see Extended Data Fig. 7a) steady-state abundances in HeLa cells reported as fold of the native  $\beta$ -globin readthrough control (RTC). **b**, Rabbit  $\beta$ -globin reporter abundances in cells transfected with siRNAs directed against human *UPF1* or *UPF2* compared to cells transfected with scrambled siRNA. **c**, HeLa Tzm cells were mock-transfected, transfected with scrambled siRNA, human *SMG1* siRNA (5–15 nM), miR-1224 precursor (5 and 20 nM), *AGO1* or *DGC8* siRNAs (10 nM each). The effects of transfected RNA species on *CCR5* mRNA steady-state abundance were assayed by qRT-PCR. **d**, Quantitative sandwich enzyme-linked immunosorbent assay (ELISA) of samples from **c**. Cell lysates (16  $\mu$ g protein per sample) were assayed and total amounts of *CCR5* protein were determined relative to standards. **e**, Computationally identified putative -1 PRF signals assayed in HeLa cells transfected with an siRNA directed against *AGO1* or a scrambled siRNA control. Numbers in human *IL8R $\alpha$*  denote the nucleotide positions of the beginning of the slippery sites in the native mRNA. **a–c**,  $n = 9$  (three times on three independent biological replicates). **d**,  $n = 8$  (quadruplicate assays of two independent biological replicates). Error bars denote standard error. \* $P < 0.05$ , \*\* $P < 0.01$  (Student's two-tailed  $t$ -test).

destabilizing activity being independent of -1 PRF-directed NMD. A transcriptional arrest time course experiment showed that the *CCR5* -1 PRF signal rendered the reporter mRNA a direct substrate for NMD: its half-life was reduced to about 180 min whereas NMD ablation increased the half-life to about 380 min (Extended Data Fig. 7b).

### NMD and miR-1224 affect CCR5 expression

In HeLa Tzm-BL cells the abundances of both *CCR5* mRNA and *CCR5* protein increased proportionally to the extent of NMD abrogation (Fig. 5c, d, siRNA\_SMG1). Conversely, addition of miR-1224 decreased both *CCR5* mRNA and *CCR5* protein abundance. Abrogation of miRNA

processing by siRNA knockdown of *AGO1* or *DGC8* resulted in increased abundance mRNA and protein, consistent with inhibition of *CCR5*-mediated -1 PRF under these conditions. Transcriptional arrest time-course experiments demonstrated that the *CCR5* mRNA is a direct substrate for NMD (Extended Data Fig. 7c). Whereas miR-1224 decreased the abundance of *CCR5* mRNA, this effect was abrogated by addition of an anti-miR-1224 antagomir, but antagomir alone had no effect (Extended Data Fig. 7d). siRNA knockdown of *SMG1* was epistatic to miR-1224, consistent with the mRNA destabilization activity of the miRNA being NMD-dependent. This is also consistent with findings that human *UPF1* may participate in RNA silencing<sup>20</sup>, with the caveat that miR-1224 may also promote mRNA degradation by a mechanism that is independent of the NMD machinery, for example, No-go decay. Combinations involving human *SMG1* siRNA knockdown plus miR-1224, human *SMG1* siRNA knockdown plus the antagomir, or all three together were also supportive of this model.

### -1 PRF and interleukin receptor mRNAs

To our knowledge before the current study, only three -1 PRF signals were known in mammalian genomes, all thought to be remnants of ancient retroviral insertional events<sup>21–23</sup>. Potential -1 PRF signals in seven additional interleukin receptor subunit mRNAs were assayed in the presence of either a scrambled siRNA control or an siRNA targeting argonaute 1. Efficient -1 PRF (>1%) was elicited by six of these (Fig. 5e). siRNA knockdown of argonaute 1 stimulated -1 PRF in some cases and inhibited it in others, consistent with sequence-specific regulation of -1 PRF by miRNAs. Ribosome profiling data<sup>10</sup> revealed ribosomes paused and directed to new reading frames at three of these signals (Extended Data Fig. 8). Single nucleotide polymorphisms (SNPs) capable of disrupting frameshifting activity were identified in all these -1 PRF signals (Extended Data Fig. 8). These may account for disease phenotypes associated with SNPs that do not alter the primary amino acid sequences of their encoded proteins.

To summarize, precise regulation of -1 PRF is accomplished by sequence-specific interactions between individual -1 PRF signals and naturally occurring miRNAs. That global ablation of miRNA processing differentially affected -1 PRF promoted by many different signals suggests that miRNA-mediated control of -1 PRF is the biologically significant norm. This confers sequence specificity, and is energetically less expensive than producing new, or modifying pre-existing ribosomes. It may also enable rapid regulation of -1 PRF on specific mRNAs within individual cells or intracellular compartments. This solves the central question, unanswered until now, of how -1 PRF may be regulated in a sequence-specific manner, and suggests a novel mode through which -1 PRF signals may be targeted for therapeutic intervention. To our knowledge, this is also one of the few demonstrations of an miRNA affecting the expression of a cellular gene through an interaction with its ORF. The discovery of -1 PRF signals in the mRNAs encoding cytokine receptors has a potentially profound impact on our understanding of immune homeostasis. Although a robust immune response is critical for limiting and preventing infection, left uncontrolled, it can rapidly result in pathology and death. Despite a large body of literature describing how expression of small-peptide mediators of the immune response are regulated at the level of mRNA stability, this only provides a global mechanism of immune regulation by controlling production of effector molecules. In contrast, the ability to control expression of cytokine receptors through -1 PRF induced NMD, and how rates of -1 PRF in turn may be controlled by miRNAs, represents a way for individual recipient cells to modulate their responses to cytokines; this would provide the means to fine-tune immune responses, and suggests a novel molecular mechanism underlying immune desensitization. The studies described here also have consequences for directing antiviral efforts. RNA viruses such as retroviruses, coronaviruses, alphaviruses and totiviruses require extremely stringent levels of -1 PRF for their propagation<sup>1</sup>. We suggest that their -1 PRF promoting structural elements may have evolved in two different ways so as to ensure set



rates of  $-1$  PRF. First, either their  $-1$  PRF stimulatory elements should not interact with any ncRNAs present in the cells in which they replicate, or their  $-1$  PRF signals may have evolved in the presence of trans-acting ncRNAs specific to their host cells. If the latter is true, as suggested by stimulation of HIV-1-promoted  $-1$  PRF in response to siRNA knock-down of argonaute, this may define a new parameter governing host cell permissiveness, presenting a novel therapeutic targeting opportunity. Thus, while the discovery of operational  $-1$  PRF signals in cellular mRNAs suggests that global targeting of  $-1$  PRF may not be the wisest approach, discovery and subsequent targeting of specific cellular miRNAs required by viruses to ensure proper rates of  $-1$  PRF may present a more narrowly targeted therapeutic option.

## METHODS SUMMARY

HeLa, HeLa Tet-Off, HeLa TZM-BL, CHO and Vero cells were cultured according to suppliers instructions. Insertions were amplified using PCR and ligated into appropriate backbone plasmids.  $-1$  PRF was assayed in live cells using dual-luciferase assays. Liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis of affinity purified CCR5/ $\beta$ -galactosidase fusion protein digested with Asp-N was performed at the University of Maryland Proteomics Core Facility. Data generated by chemical modification assays were used for *in silico* three-dimensional modelling. Sandwich enzyme-linked immunosorbent assay (ELISA) was used to monitor CCR5 protein expression, and quantitative PCR with reverse transcription (qRT-PCR) analyses were used to monitor mRNA steady-state abundance and half-lives. *In vivo* affinity capture used a double stranded miR-1224-5p RNA containing a sense strand 5' biotin modification and mismatch transfected into HeLa and HeLa TZM-BL cells. Electrophoretic mobility shift assays used a synthetic hsa-miR-1224-5p and transcripts harbouring the CCR5 or HIV-1  $-1$  PRF signals.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 February 2012; accepted 29 April 2014.**

**Published online 9 July 2014.**

- Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip. Rev. RNA* **3**, 661–673 (2012).
- Advani, V. M., Belew, A. T. & Dinman, J. D. Yeast telomere maintenance is globally controlled by programmed ribosomal frameshifting and the nonsense-mediated mRNA decay pathway. *Translation* **3**, e24418 (2013).
- Sulima, S. O. *et al.* Bypass of the pre-60S ribosomal quality control as a pathway to oncogenesis. *Proc. Natl Acad. Sci. USA* **111**, 5640–5645 (2014).
- Dinman, J. D. & O'Connor, M. in *Recoding Expansion of Decoding Rules Enriches Gene Expression*. (Atkins, J. F. & Gesteland, R. F.) 321–344 (Springer, 2010).
- Hekman, K. E. *et al.* A conserved eEF2 coding variant in SCA26 leads to loss of translational fidelity and increased susceptibility to proteostatic insult. *Hum. Mol. Genet.* **21**, 5472–5483 (2012).
- De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in *CNOT3* and ribosomal genes *RPL5* and *RPL10* in T-cell acute lymphoblastic leukemia. *Nature Genet.* **45**, 186–190 (2013).
- Jack, K. *et al.* rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Mol. Cell* **44**, 660–666 (2011).
- Dragic, T. *et al.* HIV-1 entry into CD4<sup>+</sup> cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667–673 (1996).
- Belew, A. T., Hepler, N. L., Jacobs, J. L. & Dinman, J. D. PRFdb: a database of computationally predicted eukaryotic programmed  $-1$  ribosomal frameshift signals. *BMC Genomics* **9**, 339 (2008).
- Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
- de Messieres, M. *et al.* Single-molecule measurements of the CCR5 mRNA unfolding pathways. *Biophys. J.* **106**, 244–252 (2014).
- Staple, D. W. & Butcher, S. E. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J. Mol. Biol.* **349**, 1011–1023 (2005).
- Houck-Loomis, B. *et al.* An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* **480**, 561–564 (2011).
- Ritchie, D. B., Foster, D. A. N. & Woodside, M. T. Programmed  $-1$  frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl Acad. Sci. USA* **109**, 16167–16172 (2012).
- Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J. Virol.* **72**, 2855–2864 (1998).
- Plant, E. P. & Dinman, J. D. Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res.* **33**, 1825–1833 (2005).
- Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11–23 (2001).
- Belew, A. T., Advani, V. M. & Dinman, J. D. Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. *Nucleic Acids Res.* **39**, 2799–2808 (2011).
- Djuranovic, S., Nahvi, A. & Green, R. A parsimonious model for gene regulation by miRNAs. *Science* **331**, 550–553 (2011).
- Jin, H. *et al.* Human UPF1 participates in small RNA-induced mRNA downregulation. *Mol. Cell Biol.* **29**, 5789–5799 (2009).
- Manktelow, E., Shigemoto, K. & Brierley, I. Characterization of the frameshift signal of *Edr*, a mammalian example of programmed  $-1$  ribosomal frameshifting. *Nucleic Acids Res.* **33**, 1553–1563 (2005).
- Wills, N. M., Moore, B., Hammer, A., Gesteland, R. F. & Atkins, J. F. A functional  $-1$  ribosomal frameshift signal in the human paraneoplastic *Ma3* gene. *J. Biol. Chem.* **281**, 7082–7088 (2006).
- Clark, M. B. *et al.* Mammalian gene *PEG10* expresses two reading frames by high efficiency  $-1$  frameshifting in embryonic-associated tissues. *J. Biol. Chem.* **282**, 37359–37369 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by grants to J.D.D. from the National Institutes of Health (5 R01GM058859, R21GM068123) and from the National Science Foundation (MCB-0084559). A.T.B. was supported by NIH/NIAD T32 AI051967, and a University of Maryland College of CMNS Hockmeyer Fellowship. S.O.S. was supported by NIH/NIGMS T32GM080201. This publication has also been funded in part with federal funds from the Frederick National Laboratory for Cancer Research, NIH, under Contract HHSN261200800001E to W.K.K. This research was supported in part by the Intramural Research Program of the National Institutes of Health, Center for Cancer Research to B.A.S. The content of this publication does not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

**Author Contributions** A.T.B. designed and performed experiments (Figs 1a, 3a, b, 4b, c, 5a, b, e and Extended Data Figs 1a, 2b, 4b, c, 5b–d and 6), analysed data, and helped in editing the manuscript. A.M. designed and performed experiments and analysed data (Fig. 4a, d and Extended Data Figs 2c, d, 3b, c, 5 and 6). S.M. made constructs and performed the experiment shown in Fig. 1b, and helped with experiments. V.M.A. designed, performed and analysed the data collected in the experiments shown in Figs 3c, 5c, d, and assisted with data shown in Figs 3b, 5a, b, and Extended Data Fig. 5d. S.O.S. assisted with sample preparation and data collection. W.K.K. and B.A.S. performed all of the molecular dynamics simulations and molecular modelling. J.D.D. conceived and directed the project, designed experiments and wrote the manuscript.

**Author Information** The sequences determined in this work are deposited in the NCBI Nucleotide database; a list is available as Supplementary Information. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.D. (dinman@umd.edu).

# Crystal structure of a human GABA<sub>A</sub> receptor

Paul S. Miller<sup>1</sup> & A. Radu Aricescu<sup>1</sup>

**Type-A  $\gamma$ -aminobutyric acid receptors (GABA<sub>A</sub>Rs) are the principal mediators of rapid inhibitory synaptic transmission in the human brain. A decline in GABA<sub>A</sub>R signalling triggers hyperactive neurological disorders such as insomnia, anxiety and epilepsy. Here we present the first three-dimensional structure of a GABA<sub>A</sub>R, the human  $\beta$ 3 homopentamer, at 3 Å resolution. This structure reveals architectural elements unique to eukaryotic Cys-loop receptors, explains the mechanistic consequences of multiple human disease mutations and shows an unexpected structural role for a conserved N-linked glycan. The receptor was crystallized bound to a previously unknown agonist, benzamidine, opening a new avenue for the rational design of GABA<sub>A</sub>R modulators. The channel region forms a closed gate at the base of the pore, representative of a desensitized state. These results offer new insights into the signalling mechanisms of pentameric ligand-gated ion channels and enhance current understanding of GABAergic neurotransmission.**

In response to binding the neurotransmitter GABA, released at inhibitory synapses, GABA<sub>A</sub>R chloride channels open and depress neuronal excitability in the adult central nervous system<sup>1</sup>. GABA<sub>A</sub>Rs belong to a superfamily of pentameric ligand-gated ion channels (pLGICs) known as the Cys-loop receptors, which includes the cation-selective nicotinic acetylcholine receptors (nAChRs) and serotonin type 3 receptors (5HT<sub>3</sub>Rs), as well as anion-selective glycine receptors (GlyRs)<sup>2</sup>. These molecules assemble as pentamers from a variety of subunits. In human GABA<sub>A</sub>Rs these subunits are encoded by 19 different genes:  $\alpha$ 1–6,  $\beta$ 1–3,  $\gamma$ 1–3,  $\delta$ ,  $\epsilon$ ,  $\theta$ ,  $\pi$  and  $\rho$ 1–3 (ref. 3). Most physiological heteromeric formats are thought to include two  $\alpha$ , two  $\beta$  and one other, most frequently a  $\gamma$  subunit<sup>4</sup>.  $\beta$ 3 subunits can also efficiently assemble into functional homomeric channels, and although they have yet to be identified as discrete populations in the brain, they serve as meaningful models for the heteromeric receptors<sup>5</sup>. Each subunit contributes an extracellular domain of 200–250 amino acids, an  $\alpha$ -helical M1–M4 transmembrane bundle and an M3–M4 intracellular loop of 85–255 residues<sup>6</sup>. Neurotransmitter molecules bind at extracellular pockets between subunits to induce a conformational switch that crosses to the transmembrane region to open the ion channel<sup>7</sup>. GABA<sub>A</sub>Rs are the targets of a wide range of drugs including benzodiazepines<sup>8</sup>, used in the treatment of epilepsy, insomnia, anxiety and panic disorder, and the intravenous general anaesthetics propofol and etomidate<sup>5,9</sup>. GABA<sub>A</sub>Rs also mediate alcohol inebriation<sup>10</sup> and are targets for endogenous modulators such as neurosteroids<sup>11</sup>.

In the absence of GABA<sub>A</sub>R structural information, insights relied on analogies with related proteins. The soluble acetylcholine binding protein (AChBP) provided the first high-resolution model for the extracellular region<sup>12</sup>. Ground-breaking electron microscopic studies gradually led to an atomic model (at 4 Å) of a complete heteromeric nAChR from the *Torpedo marmorata* electric organ<sup>13,14</sup>, and a general framework for molecular understanding of pLGICs. Subsequently, crystal structures of two bacterial homologues, ELIC<sup>15</sup> and GLIC<sup>16</sup>, as well as the first structure of an anion-selective Cys-loop receptor, the *Caenorhabditis elegans* glutamate-gated chloride channel  $\alpha$  homopentamer (GluCl $\alpha$ )<sup>17</sup>, were reported, providing insights into potential mechanisms of interaction with orthosteric ligands and allosteric modulators<sup>17,18</sup>. Nevertheless, owing to limited sequence identities with GABA<sub>A</sub>R subunits, these models alone cannot adequately explain their rules of assembly, ligand binding

and modulation, gating mechanism, or the consequences of numerous human mutations linked to epilepsy and insomnia. Aiming to address these unknowns, we report here the crystal structure of a human GABA<sub>A</sub>R, the  $\beta$ 3 homopentamer.

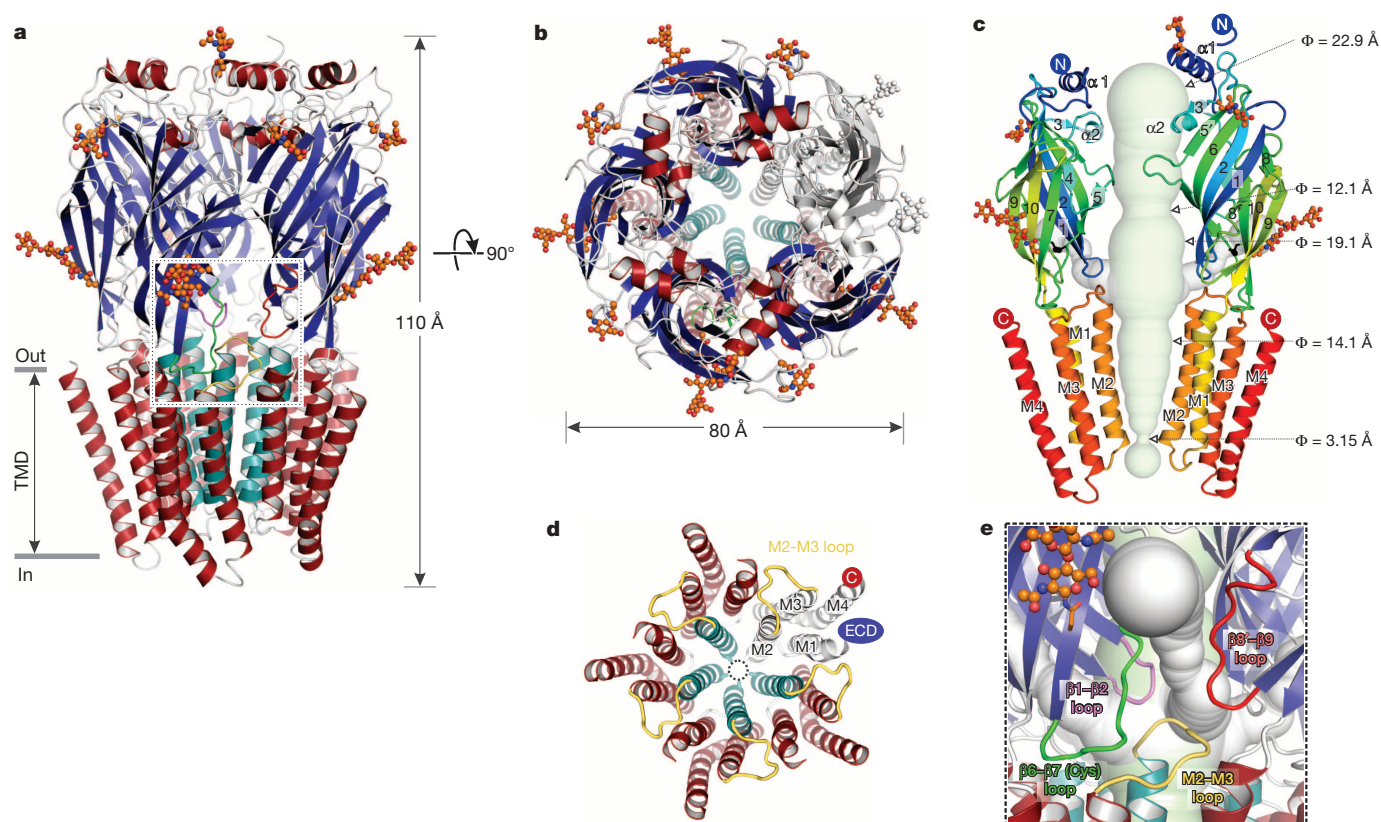
## Architecture of the GABA<sub>A</sub>R $\beta$ 3 homopentamer

Crystallization of human GABA<sub>A</sub>R  $\beta$ 3 required truncation of the intracellular loop between transmembrane helices 3 and 4 (M3–M4). Residues from Gly 308 to Asn 421 were substituted by a linker sequence, SQPARAA<sup>19</sup>, to give the construct GABA<sub>A</sub>R- $\beta$ 3<sub>cryst</sub> (residue numbering used here corresponds to the mature isoform 1, that is, Gln 26 in UniProt entry P28472 is Gln 1 in GABA<sub>A</sub>R- $\beta$ 3<sub>cryst</sub>). GABA<sub>A</sub>R- $\beta$ 3<sub>cryst</sub> solubilized in detergent retained the ability to bind agonists (GABA and histamine), channel blockers (fipronil and picrotoxin) and the anaesthetic etomidate (Extended Data Fig. 1a–e). Furthermore, in patch-clamped HEK293T cells expressing GABA<sub>A</sub>R- $\beta$ 3<sub>cryst</sub>, application of the GABA<sub>A</sub>R  $\beta$ 3 agonists histamine or propofol induced inward currents that were inhibited by the channel blockers fipronil and picrotoxin (Extended Data Fig. 1f, g).

We crystallized and determined the structure of GABA<sub>A</sub>R- $\beta$ 3<sub>cryst</sub> at 3 Å resolution (Extended Data Table 1). Viewed perpendicular to the central five-fold pseudo-symmetry axis, the receptor approximates a cylinder 110 Å in height, with diameter ranging from 60 to 80 Å (Fig. 1a), spanning the plasma membrane and protruding ~65 Å into the extracellular space. Viewed along the same pseudo-symmetry axis from the extracellular side, the pentamer has a toroidal (doughnut-like) profile, surrounded by 15 (three per subunit) N-linked glycans (Fig. 1b). Each extracellular domain (ECD) comprises an amino-terminal  $\alpha$ -helix ( $\alpha$ 1) followed by ten  $\beta$ -strands folded into a curled  $\beta$ -sandwich, topologically similar to other family members characterized to date (Fig. 1c and Extended Data Figs 2 and 3). A second  $\alpha$ -helix ( $\alpha$ 2), between  $\beta$ -strands 3 and 4, is located under the  $\alpha$ 1 helix (Fig. 1c). Four additional helices (M1–M4) from each subunit come together to create the pentameric transmembrane domain (TMD), with M2 segments lining a pore that tapers as it traverses towards the intracellular side of the membrane (Fig. 1b–d). On the extracellular side, water and solutes access the pore from a vestibule surrounded by the ECDs and from lateral tunnels situated between the ECDs that are lined by an excess of negatively charged

<sup>1</sup>Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.





**Figure 1 | Architecture of GABA<sub>A</sub>-R-β<sub>3</sub><sub>cryst</sub>.** **a**, GABA<sub>A</sub>-R-β<sub>3</sub><sub>cryst</sub> viewed parallel to the plasma membrane ( $\alpha$ -helices red, except the pore-lining M2 shown in teal;  $\beta$ -strands blue; loops grey). N-linked glycans are shown in orange ball-and-stick representation. **b**, View from the extracellular space (synaptic cleft) down the five-fold pseudo-symmetry axis, with a single subunit coloured in grey. **c**, Two subunits, rainbow coloured from blue N terminus to red C terminus, illustrating secondary structure nomenclature.

A water-filled ECD vestibule and TMD pore shown in light green (diameter ( $\Phi$ ) indicated periodically) runs through the five-fold pseudo-symmetry axis of the pentamer, joined by lateral tunnels coming from between each of the subunit ECDs (only two shown for clarity, in grey). **d**, The pentameric transmembrane region, to illustrate the arrangement of helices M1–M4 and the M2–M3 loop (yellow). **e**, View of a lateral tunnel running between subunits into the central vestibule.

groups, suggesting putative cation modulation sites (Fig. 1c, e and Extended Data Fig. 4a, d, g, h).

A positively charged ring halfway down the vestibule hosts putative anion binding sites at each inter-subunit interface (Extended Data Fig. 4d), revealed by peaks (visible to an  $\sim 6\sigma$  level) in  $F_o - F_c$  electron density maps calculated following refinement without anions modelled (Extended Data Fig. 5a). Attempts to confirm the nature of these peaks by soaking heavy atom anions (for example, iodide) into the crystals and analysis of anomalous electron density maps were inconclusive. However, placement of chloride ions followed by refinement satisfactorily accounts for these electron density features (Extended Data Fig. 5a–c). This positively charged ring corresponds to the previously proposed ion selectivity filter in the vestibule of Cys-loop receptors<sup>20</sup>. Furthermore, these putative chloride sites are in spatial proximity (but structurally non-equivalent) to the ‘anion site 1’ reported in the bacterial channel GLIC<sup>21</sup>, and thus chloride ions might be important stabilizers of pLGIC assembly.

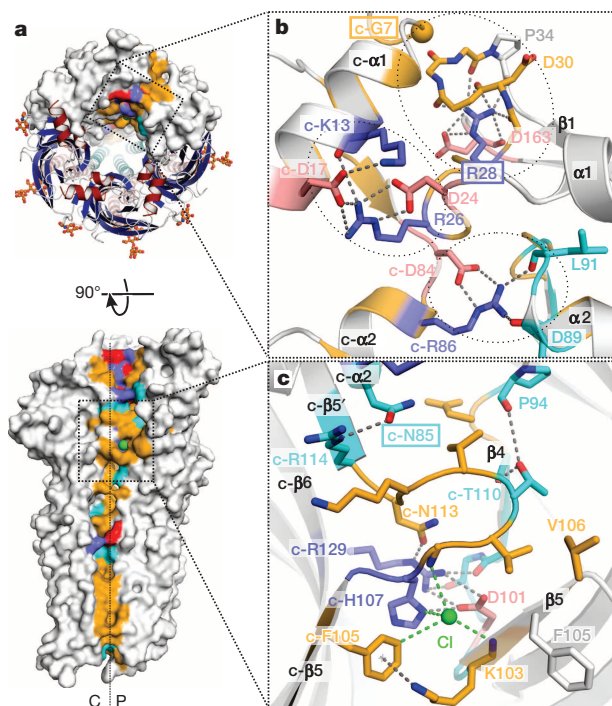
At the TMD level, two large non-overlapping pockets are located near residues previously inferred to bind the intravenous anaesthetics etomidate and propofol<sup>5,9</sup> (Extended Data Fig. 6a–c). The putative propofol-binding pocket is structurally distinct from the one identified in the bacterial channel GLIC<sup>18</sup>. The binding and transduction modes of propofol on GABA<sub>A</sub>-R-β<sub>3</sub><sub>cryst</sub> and GLIC are therefore unrelated, as are its relative potencies. Propofol potentiates and activates GABA<sub>A</sub>Rs but inhibits GLIC<sup>5,18</sup>.

## Determinants of assembly

The 19 different GABA<sub>A</sub>R subunits obey stringent rules for assembly specificity<sup>4</sup>, the molecular determinants of which are poorly understood. A comparison of inter-subunit interfaces from pLGIC structures available

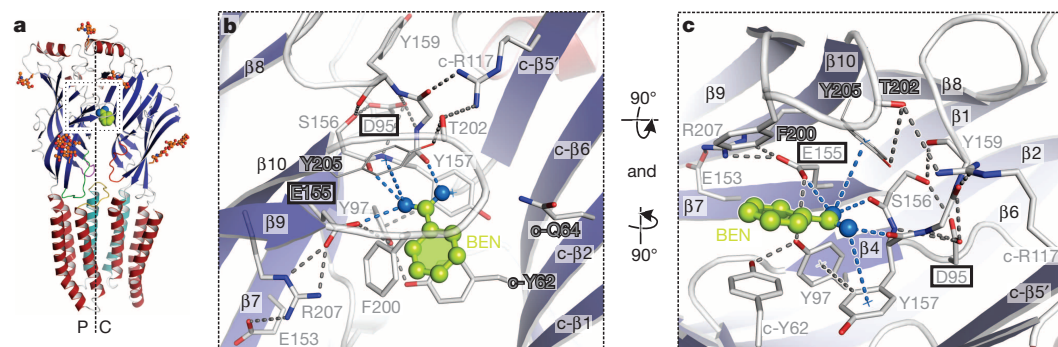
to date reveals considerable differences in the geometry and thermodynamics of complex formation, with GABA<sub>A</sub>-R-β<sub>3</sub><sub>cryst</sub> subunits forming the most extensive, energetically favourable interactions (Extended Data Fig. 7b). These occur in particular between ECDs, through patchworks of hydrogen bonds, salt bridges and van der Waals contacts (Fig. 2a–c and Extended Data Fig. 7a–e). Within the upper portions of the ECDs, between the  $\alpha 1$  helices of adjacent subunits, the side chains of Arg 26 and Asp 17 form a salt-bridge network extending to Asp 24 and Lys 13 (Fig. 2b), an interaction specific to  $\beta$ – $\beta$  and  $\alpha$ – $\beta$  subunits in human GABA<sub>A</sub>Rs (Extended Data Fig. 3). Between the  $\alpha 2$  helices, the inter-subunit interface is stabilized by a network of hydrogen bonds and salt bridges surrounding Arg 86, specific to the GABA<sub>A</sub>R  $\beta$  and GlyR subunits (Fig. 2b and Extended Data Fig. 3). The loop connecting the  $\alpha 1$  helix with the  $\beta 1$  strand also bridges this interface and mutations in it (or its proximity) are associated with childhood absence epilepsy and febrile seizures<sup>22,23</sup> (Extended Data Fig. 3). The GABA<sub>A</sub>R  $\beta 3$  Gly7Arg mutation that excludes  $\gamma 2$  subunits from heteromeric GABA<sub>A</sub>Rs<sup>22</sup> opposes and is likely to perturb the  $\alpha 1$ – $\beta 1$  loop conformation (Fig. 2b). The GABA<sub>A</sub>R  $\gamma 2$  Arg43Gln mutation (equivalent to Arg28Gln in GABA<sub>A</sub>-R-β<sub>3</sub><sub>cryst</sub>), that also excludes  $\gamma 2$  subunits<sup>23</sup>, resides within the  $\alpha 1$ – $\beta 1$  loop and is predicted to disrupt its backbone organization (Fig. 2b).

Inter-subunit contacts between the central portions of ECDs involve the  $\beta 4$ ,  $\beta 5$ ,  $\beta 5'$  and  $\beta 6$  strands and flanking loops. The  $\beta 5$ – $\beta 5'$  loop is extended in GABA<sub>A</sub>-R-β<sub>3</sub><sub>cryst</sub> and GluCl $\alpha$  compared to nAChRs and AChBP (Extended Data Fig. 2d), and protrudes into the neighbouring subunit (Fig. 2c). The  $\beta 5$ – $\beta 5'$  loop His 107, strictly conserved in GABA<sub>A</sub> and Gly receptors, and main-chain amino groups in positions 104 and 108 coordinate putative chloride anions across this boundary (within the positively charged ring lining the vestibule discussed earlier; Fig. 2c



**Figure 2 | Assembly interactions in GABA<sub>A</sub>β<sub>3</sub>-cryst.** **a**, Top-down view of the GABA<sub>A</sub>β<sub>3</sub>-cryst pentamer and side-on view of two neighbouring subunits from the vestibule, highlighting the nature of inter-subunit contacts between the principal (P) face of one subunit and complementary (C) face (residues marked by 'c-') of the next. Salt-bridging residues are coloured purple and red, those forming putative hydrogen bonds in cyan, and residues forming van der Waals contacts in orange. **b**, Upper ECD close-up showing the inter-subunit α1-β1 loop (upper dotted oval), the inter-subunit salt bridges connecting α1 helices (middle dotted oval) and inter-subunit α2-α2 interactions (lower dotted oval). Boxed residue labels correspond to disease mutations discussed in main text. **c**, The ECD anion-binding site and surrounding inter-subunit interface (chloride shown as a green sphere). Grey dashed lines indicate putative salt bridges and hydrogen bonds, green dashes indicate chloride coordination.

and Extended Data Fig. 5c). Further interface contacts include a salt-bridge network centred on Asp 101 and Arg 129, and a potential cation-π interaction between Lys 103 and Phe 105, both likely to influence assembly specificity based on residue conservation patterns (Fig. 2c and Extended Data Fig. 3). Mutations within this ECD interface are linked to epileptic encephalopathies, for example Asn85Asp<sup>24</sup>, which will remodel interactions with basic residues adjoining the β5-β5' loop (Fig. 2c).



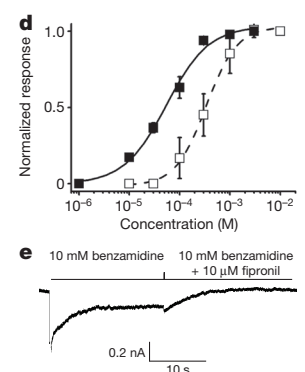
**Figure 3 | Neurotransmitter pocket occupied by the agonist benzamidine.** **a**, Benzamidine (green/blue spheres) bound in the neurotransmitter pocket. Of note, the β8-β8' loop, known as loop F (Arg 169–Ala 174; running into the red loop) in heteromeric GABA<sub>A</sub>Rs, does not contribute to the GABA<sub>A</sub>β<sub>3</sub>-cryst orthosteric site. **b**, **c**, Benzamidine binding mode. Complementary face residues marked by 'c-'. Grey dashed lines indicate putative hydrogen bonds, salt bridges and cation-π interactions; blue dashed lines indicate the coordination

## Neurotransmitter pocket and binding of a novel agonist

The neurotransmitter-binding pocket of GABA<sub>A</sub>Rs is located between ECDs. It comprises the β4 strand and adjacent residues (Asp 95–Leu 99), part of the β7–β8 loop (Glu 155–Tyr 159) and the β9–β10 loop (Phe 200–Tyr 205) from the 'principal' face, also known as loops A–C, respectively. The 'complementary' face in GABA<sub>A</sub>β<sub>3</sub>-cryst comprises a segment of the β2 strand (Tyr 62–Gln 64) and the β6 strand (Leu 125–Arg 129), also known as loops D and E, respectively, which come from equivalent motifs in α subunits in heteromeric GABA<sub>A</sub>Rs (Fig. 3a–c and Extended Data Fig. 3)<sup>25</sup>. The β9–β10 loop adopts a 'closed' conformation over the site, consistent with an agonist-bound conformation<sup>17,26–28</sup> (Extended Data Fig. 8). Closure is stabilized by salt bridges between the side chains of Arg 207, Glu 153 and Glu 155 (Fig. 3b, c), residues previously implicated in GABA binding and channel activation<sup>29–31</sup>. Analogous interactions are required for ligand binding and activation of nAChRs<sup>12,32</sup>.

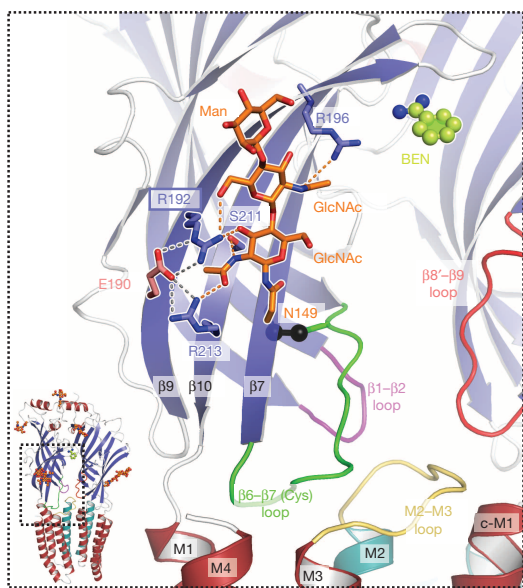
We observed large positive peaks in the  $F_o - F_c$  electron density map in all five neurotransmitter-binding sites, which could be accounted for by benzamidine molecules, an additive that helped us obtain high-resolution diffracting crystals (Extended Data Fig. 5d–f). To our knowledge, benzamidine or its derivatives have not been previously reported to act as GABA<sub>A</sub> ligands. We recorded GABA<sub>A</sub>β<sub>3</sub>-cryst currents in HEK293S-GnTI<sup>-</sup> (deficient in N-acetylglucosaminyltransferase I activity) cells, and found benzamidine to behave as an agonist capable of inducing desensitization (half-maximal effective concentration ( $EC_{50}$ ) =  $61 \pm 12 \mu\text{M}$ ;  $n = 4$ ; Fig. 3d, e). Thermostabilization of GABA<sub>A</sub>β<sub>3</sub>-cryst in detergent micelles by benzamidine revealed a sensitivity similar to histamine (benzamidine  $EC_{50} = 370 \pm 180 \mu\text{M}$ ; histamine  $EC_{50} = 400 \pm 150 \mu\text{M}$ ;  $n = 3$ ; Fig. 3d and Extended Data Fig. 1c). Classically, benzamidine is known as a highly potent serine-protease inhibitor, derivatives of which are in clinical trials for prevention of blood clotting<sup>33</sup>. However, given that GABA<sub>A</sub>β<sub>3</sub> agonists such as histamine act instead as potentiators of heteromeric GABA<sub>A</sub>Rs, characterization of benzamidine derivatives as positive modulators may offer new opportunities in drug development.

The benzamidine benzyl ring is stacked between the side chains of Phe 200 and Tyr 62, whereas its amidinium group forms hydrogen bonds with the Glu 155 side chain and backbone carbonyls of Ser 156 and Tyr 157, and putative cation-π interactions with the Tyr 157 and Tyr 205 aromatic rings (Fig. 3b, c). This binding mode is reminiscent of the principal face observed in GluClα and AChBP (Extended Data Fig. 8a, b). Two epileptic encephalopathies are linked to mutations in this β3 subunit region, Asp95Asn and Glu155Gly<sup>24</sup>. Loss of Asp 95, which hydrogen bonds Ser 156 and Tyr 157, will probably destabilize the β7–β8 loop conformation, whereas Glu155Gly will impair binding of amino-group-containing ligands (Fig. 3b, c and Supplementary Discussion).



**e**, Electrophysiological response to 10 mM benzamidine and block by 10 μM fipronil (an alternative blocker, picrotoxin, has a similar effect on GABA<sub>A</sub>β<sub>3</sub>-cryst currents; Extended Data Fig. 1f).





**Figure 4 | A conserved glycosylation site interacts with  $\beta 9$ – $\beta 10$  loop residues.** Close-up of the N-linked glycosylation site 3, attached to the  $\beta 7$ -strand that supports the  $\beta 6$ – $\beta 7$  (Cys) loop (Cys-bridge shown in black spheres) and its interactions with surrounding residues from the  $\beta 9$ – $\beta 10$  agonist-binding loop (indicated by dashed orange lines; grey dashed lines highlight additional putative salt bridges). Boxed residue labels correspond to disease mutations discussed in main text.

### The impact of N-linked glycosylation

Two of the three N-linked glycosylation sites present in each subunit (Asn 8, only visible in chain A, and Asn 80) are occupied by residual GlcNAc moieties following endoglycosidase F1 treatment. The third site, Asn 149 on the  $\beta 7$  strand (Fig. 4), is conserved in almost all GABA<sub>A</sub>  $\beta$ , nAChR and 5HT<sub>3</sub>R subunits (Extended Data Fig. 3) and was resistant to enzymatic digestion (Extended Data Fig. 5g–i). This glycan extends along the  $\beta 9$  and  $\beta 10$  strands that support the agonist-binding ‘loop C’, providing further contacts between these units and the  $\beta 7$  strand, which may facilitate the ECD–TMD signal transduction. Substitution of Asn 149 in GABA<sub>A</sub>  $\beta 2$  reduces sensitivity to GABA<sup>34</sup>, and in GlyRs, Zn<sup>2+</sup> coordination between the same three  $\beta$ -strands potentiates agonist sensitivity<sup>35</sup>. Furthermore, mutation of Arg 192 at the core of this network of glycan interactions is associated with chronic insomnia and increases the inactivation rate of GABA<sub>A</sub>Rs<sup>36</sup>. Importantly, the equivalent glycan in nAChR  $\alpha 1$  also contacts the outer face of the ECD, and its enzymatic cleavage decreases nAChR function<sup>37</sup>.

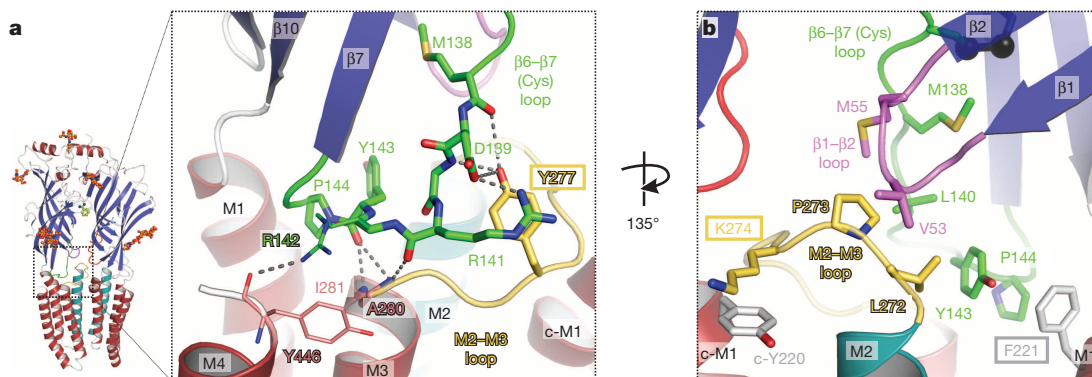
### Structural coupling between the ECD and TMD

For a neurotransmitter binding event to transmit to the channel gate, a signal must be transduced across the ECD–TMD interface. The inner and outer  $\beta$ -sheets of each GABA<sub>A</sub>  $\beta 3_{\text{cryst}}$  ECD come together at the base through a conserved salt-bridge between  $\beta 1$ – $\beta 2$  loop Glu 52 and  $\beta 10$ -strand Arg 216 (refs 14, 38) (Extended Data Fig. 7e). Beneath this, in each subunit the ECD–TMD interface consists of two clusters of interactions: an array of polar contacts linking the outer portion of the M2–M3 loop with the  $\beta 6$ – $\beta 7$  (‘Cys’) loop (Fig. 5a); and van der Waals contacts between residues in the inner portion of the M2–M3 loop near the pore, the  $\beta 1$ – $\beta 2$  loop and the  $\beta 6$ – $\beta 7$  loop (Fig. 5b). Central to both clusters is Pro 144 in the  $\beta 6$ – $\beta 7$  loop, conserved in all pLGICs (Extended Data Fig. 3) and adopting a *cis* conformation, which orients the neighbouring Tyr 143 backbone carbonyl downwards to engage in hydrogen bonds with M3 helix backbone amino groups (Fig. 5a). These contacts ensure that the GABA<sub>A</sub>  $\beta 3_{\text{cryst}}$  ECD–TMD interface is tightly structurally coupled, with a solvent-inaccessible surface area of 689 Å<sup>2</sup>, 100 Å<sup>2</sup> larger than observed in ELIC, but comparable to those observed in GluCl $\alpha$ , GLIC, and the three nAChR structures solved in a membrane context by electron microscopy (Extended Data Fig. 9a). Nevertheless, despite similarly strong interfaces, structural alignments of GABA<sub>A</sub>  $\beta 3_{\text{cryst}}$  with other pLGICs reveal differences in their ECD–TMD relative orientations, where rotations within a  $\sim 20^\circ$  range probably reflect the multiplicity of states (resting, activated and desensitized) that these receptors occupy<sup>39</sup> (Extended Data Fig. 9b–f).

Mutations linked to epileptic encephalopathies highlight the importance of the ECD–TMD region<sup>24,40</sup>. One such mutation, GABA<sub>A</sub>  $\beta 3$  Tyr277Cys, will reduce hydrogen-bond connectivity between the M2–M3 loop and the  $\beta 6$ – $\beta 7$  loop and disrupt a stacking interaction with the Arg 141 guanidinium group (Fig. 5a). Another mutation, GABA<sub>A</sub>  $\beta 1$  Phe246Ser (Phe 221 in  $\beta 3$ ), is located at the top of M1 and will disrupt its interaction with the critical Tyr 143/Pro 144 motif at the apex of the  $\beta 6$ – $\beta 7$  loop (Fig. 5b). A Lys289Met mutation reported in GABA<sub>A</sub>  $\gamma 2$  (Lys 274 in  $\beta 3$ ) affects a residue whose side chain, very well ordered in electron density maps, reaches across a neighbouring subunit (Fig. 5b and Extended Data Fig. 7e). This residue is conserved in GlyRs, for which mutations are linked to the rare human genetic startle disorder, hyperkplexia, and where detailed kinetic analysis of GlyR  $\alpha 1$  Lys276Glu has revealed a significantly slower gating<sup>41</sup>. Thus, on the basis of the GABA<sub>A</sub>  $\beta 3_{\text{cryst}}$  structure, it appears that this lysine facilitates the coordination of inter-subunit motions.

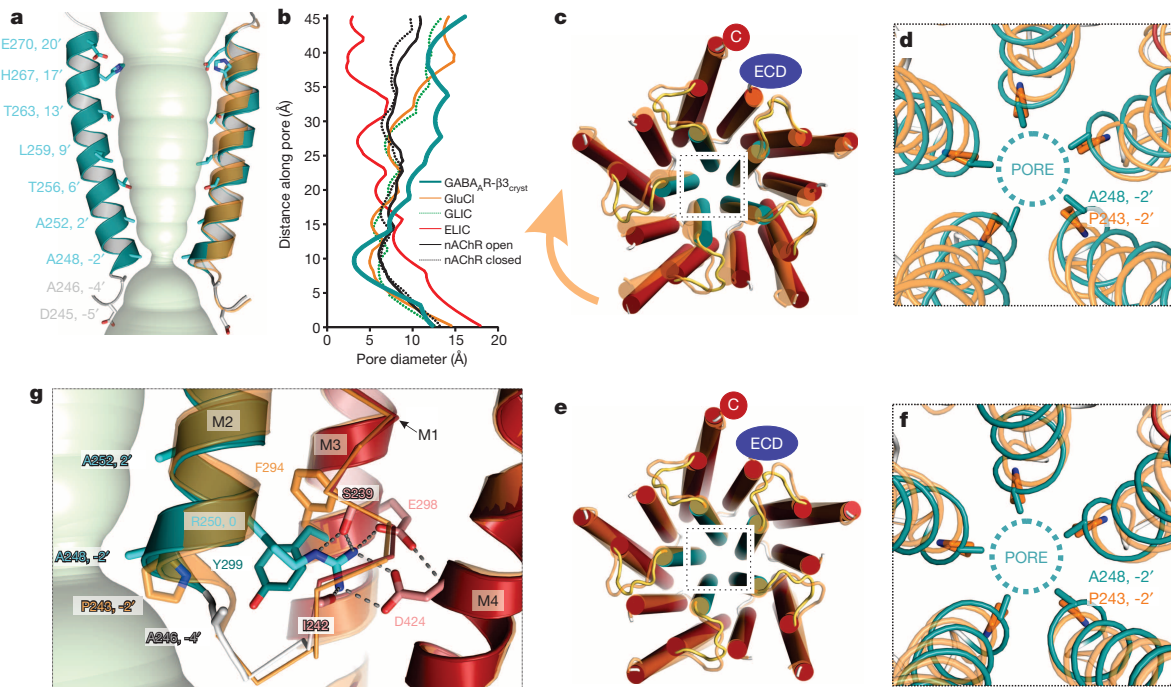
### Channel structure and the desensitization mechanism

The GABA<sub>A</sub>  $\beta 3_{\text{cryst}}$  pore is lined by five M2 helices (Fig. 1a–d), supporting ‘rings’ of residues between positions  $-5'$  and  $20'$  (Fig. 6a). In GABA<sub>A</sub>  $\beta 3_{\text{cryst}}$  the M2 helices taper inwards from the  $13'$  Thr 263



**Figure 5 | Structural coupling at the ECD–TMD interface.** **a**, Side-on view of the ECD–TMD interface, rotated  $135^\circ$  in **b**. **a**, Putative hydrogen bonds (indicated by grey dashed lines) between residues of the  $\beta 6$ – $\beta 7$  (Cys) loop, the outer portion of the M2–M3 loop and the top of M3 and M4 helices. **b**, Hydrophobic

packing in the ECD–TMD interface viewed from the pore side, involving residues from the  $\beta 6$ – $\beta 7$  loop, the  $\beta 1$ – $\beta 2$  loop, the inner portion of the M2–M3 loop and the N terminus of the M1 helix from a neighbouring subunit (c-M1). Boxed residue labels correspond to disease mutations discussed in main text.



**Figure 6 | Structure of the ion channel in a desensitized state.** **a**, Two GABA<sub>A</sub>-β3<sub>cryst</sub> M2 helices (teal), with side chains of pore-lining residues in stick representation. An equivalent GluClα (Protein Data Bank (PDB) accession 3RIF) M2 helix, in orange, illustrates its distinct flexure. **b**, Pore diameter of GABA<sub>A</sub>-β3<sub>cryst</sub> (teal) and related structures: open GluClα, open GLIC (PDB 4HFI), open nAChR (PDB 4AQ9), closed nAChR (PDB 2BG9) and closed ELIC (PDB 2VL0). **c**, Chain A superposition of pentameric GABA<sub>A</sub>-β3<sub>cryst</sub> (red/

teal) over GluClα (orange), revealing the relative rotation of transmembrane regions. **d**, The pore constriction at -2' Ala 248 in GABA<sub>A</sub>-β3<sub>cryst</sub> compared to GluClα at -2' Pro 243, using alignment in **c**. **e**, Superposition of individual GABA<sub>A</sub>-β3<sub>cryst</sub> subunit TMDs over GluClα removes the relative rotation, but the pore remains shut (**f**). **g**, GABA<sub>A</sub>-β3<sub>cryst</sub> (red/teal) showing the Tyr 299 side-chain 'pressing' M2 to constrict the channel. In GluClα (orange) Phe 294 points upwards, enabling an open pore conformation.

down to the intracellular border where the -2' Ala 248 side chains define the narrowest point, just 3.15 Å in diameter (Fig. 6a, b). This pore is too narrow to permit the passage of chloride anions (with a Pauling radius of 1.8 Å) and therefore delineates a closed gate. However, this geometry fundamentally differs from closed structures reported to date, ELIC<sup>15</sup> and *Torpedo* nAChR<sup>14</sup>, which exhibit almost vertical M2 helices and have closed gates in the extracellular portion of the pore (9' up to 20') formed by bulky hydrophobic side chains (Fig. 6b). In GABA<sub>A</sub>-β3<sub>cryst</sub> none of the M2 hydrophobic rings (1' Val, 3' Leu, 5' Ile, 8' Val, 9' Leu, 11' Met, 14' Ile and 18' Leu) line the pore. Instead, its trajectory more closely resembles the open conformations of GLIC<sup>16</sup> and GluClα<sup>17</sup>, which are narrowest at the intracellular border (Fig. 6b and Extended Data Fig. 9g).

The pore-lining residues in GABA<sub>A</sub>-β3<sub>cryst</sub> (Fig. 6a) are generally in excellent agreement with those identified by cysteine accessibility studies in the GABA<sub>A</sub> α1 subunit and by protonation of introduced charged residues in the open nAChR<sup>42,43</sup>. However, the 9' Leu residues in GABA<sub>A</sub>-β3<sub>cryst</sub> (conserved across the mammalian Cys-loop receptors) are rotated out of the pore, with side chains placed between neighbouring M2 helices. This rotation is not observed in the open-to-closed M2 flexion motion in nAChR<sup>44</sup>. Irrespectively, the M2 rotation alone cannot explain the closed state of GABA<sub>A</sub>-β3<sub>cryst</sub> because superposition of each TMD individually onto the open GluClα equivalents, to remove its influence, reveals that the GABA<sub>A</sub>-β3<sub>cryst</sub> pore would still remain closed (Fig. 6c–f). Thus, the closed state of GABA<sub>A</sub>-β3<sub>cryst</sub> is principally accounted for by a unique conformation of its M2 helix (Fig. 6a and Extended Data Fig. 9g). The expanded extracellular portion of the pore in GABA<sub>A</sub>-β3<sub>cryst</sub> is stabilized by two rings of salt bridges. The first one involves 17' His 267 and 20' Glu 270 from adjacent M2 helices (Extended Data Fig. 7e). His 267 lines both the pore and an inter-subunit cavity, coordinating Zn<sup>2+</sup> for inhibition or propofol for potentiation<sup>5,45</sup> (Extended Data Figs 6b, c and 7e). The second ring is an intra-subunit interaction between the side chains of 19' Arg 269 and M3 Asp 282, which retracts M2 against M3 (Extended Data Fig. 7e). This ring

is conserved in GABA<sub>A</sub> α-subunits and GlyR α-subunits, and substitutions of the equivalent 19' Arg 271 in GlyR α1 are the most frequent cause of hyperekplexia, with disease mutations Arg271Leu and Arg271Gln decoupling agonist binding from gating<sup>46</sup>.

The contracted intracellular portion of the pore is confined by the conformation of Tyr 299 side chains, which point towards the back of the M2 gate, compressing it shut (Fig. 6g). Aromatic Phe or Tyr residues occupy this position across all human GABA<sub>A</sub> and GlyR subtypes and in *C. elegans* GluClα (Extended Data Fig. 3). Notably, however, in the GluClα open pore conformation the equivalent Phe 294 side chain is rotated upwards, preventing such compression (Fig. 6g). During the gating process, conformation switching of side chains at this key position might account for local desensitization, and drugs that potentiate function by blocking desensitization<sup>47–49</sup> might do so by disabling this aromatic switch. A desensitized state of GABA<sub>A</sub>-β3<sub>cryst</sub> is in agreement with our electrophysiological recordings of benzamidine-induced desensitizing currents measured in HEK cells at saturating concentrations (10 mM), which approach those used in crystallization (33 mM; Fig. 3e). Furthermore, in heteromeric GABA<sub>A</sub>Rs, swapping the β-subunit intracellular border with the equivalent nAChR α7 residues renders the receptor cation selective and ablates desensitization to GABA<sup>50</sup>.

## Conclusion

Here we present the first X-ray structure of a GABA<sub>A</sub> receptor, the human β3 homopentamer, co-crystallized with a novel agonist, benzamidine. GABA<sub>A</sub>-β3<sub>cryst</sub> has a closed β9–β10 loop, being in an agonist-bound state, but the pore is shut, consistent with a desensitized conformation. To our knowledge, this is the first time when a pLGIC-desensitized state has been described crystallographically. These results shed new light on the conformational transitions that occur across pLGICs and provide a rational basis for understanding how multiple human disease mutations affect GABA<sub>A</sub> assembly, glycosylation and agonist binding, as well as the signal transduction and gating processes.



## METHODS SUMMARY

GABA<sub>A</sub>R-β<sub>3</sub><sub>cryst</sub> was expressed in HEK293S-GnTI<sup>−</sup> cells and immuno-affinity purified on Rho-1D4 antibody-coated beads that bind the carboxy-terminal 1D4 tag (TETSQVAPA). Pure GABA<sub>A</sub>R-β<sub>3</sub><sub>cryst</sub> was further isolated by size-exclusion chromatography and peak fractions were concentrated to 3 mg ml<sup>−1</sup>. Crystallization was performed by sitting-drop vapour diffusion at 4 °C with a precipitating solution containing 11.5% PEG 4000, 100 mM NaCl, 100 mM Li<sub>2</sub>SO<sub>4</sub>, 100 mM N-(2-Acetamido)iminodiacetic acid buffer, pH 6.5, and 2% (v/w) benzamidine. Crystals were cryoprotected by soaking in precipitant solution supplemented with 20% glycerol. The GABA<sub>A</sub>R-β<sub>3</sub><sub>cryst</sub> X-ray structure was solved at 3 Å resolution, using the *C. elegans* glutamate-gated chloride channel α (GluClα<sup>17</sup>) as a molecular replacement model. Phasing was followed by iterative rounds of manual model building and crystallographic refinement, leading to a complete model of the protein core. N-linked glycans, anions and benzamidine were subsequently added to the model and refined. Electrophysiological recordings were made from either transiently transfected HEK293T cells or HEK293S-GnTI<sup>−</sup> cells stably expressing GABA<sub>A</sub>R-β<sub>3</sub><sub>cryst</sub>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 December 2013; accepted 28 March 2014.

Published online 8 June; corrected online 20 August 2014 (see full-text HTML version for details).

- Rabow, L. E., Russek, S. J. & Farb, D. H. From ion currents to genomic analysis: recent advances in GABA<sub>A</sub> receptor research. *Synapse* **21**, 189–274 (1995).
- Grenningloh, G. et al. Glycine vs GABA receptors. *Nature* **330**, 25–26 (1987).
- Simon, J., Wakimoto, H., Fujita, N., Lalonde, M. & Barnard, E. A. Analysis of the set of GABA<sub>A</sub> receptor genes in the human genome. *J. Biol. Chem.* **279**, 41422–41435 (2004).
- Sigel, E. & Steinmann, M. E. Structure, function, and modulation of GABA<sub>A</sub> receptors. *J. Biol. Chem.* **287**, 40224–40231 (2012).
- Yip, G. M. et al. A propofol binding site on mammalian GABA<sub>A</sub> receptors identified by photolabeling. *Nature Chem. Biol.* **9**, 715–720 (2013).
- Karlin, A. & Akabas, M. H. Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins. *Neuron* **15**, 1231–1244 (1995).
- Miller, P. S. & Smart, T. G. Binding, activation and modulation of Cys-loop receptors. *Trends Pharmacol. Sci.* **31**, 161–174 (2010).
- Rudolph, U. & Knoflach, F. Beyond classical benzodiazepines: novel therapeutic potential of GABA<sub>A</sub> receptor subtypes. *Nature Rev. Drug Discov.* **10**, 685–697 (2011).
- Li, G. D. et al. Identification of a GABA<sub>A</sub> receptor anesthetic binding site at subunit interfaces by photolabeling with an etomidate analog. *J. Neurosci.* **26**, 11599–11605 (2006).
- Wallner, M., Hancher, H. J. & Olsen, R. W. Ethanol enhances α<sub>4</sub>β<sub>3</sub>δ and α<sub>6</sub>β<sub>3</sub>δ γ-aminobutyric acid type A receptors at low concentrations known to affect humans. *Proc. Natl Acad. Sci. USA* **100**, 15218–15223 (2003).
- Belelli, D. & Lambert, J. J. Neurosteroids: endogenous regulators of the GABA<sub>A</sub> receptor. *Nature Rev. Neurosci.* **6**, 565–575 (2005).
- Brejck, K. et al. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Hilf, R. J. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
- Bocquet, N. et al. X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* **457**, 111–114 (2009).
- Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* **474**, 54–60 (2011).
- Nury, H. et al. X-ray structures of general anaesthetics bound to a pentameric ligand-gated ion channel. *Nature* **469**, 428–431 (2011).
- Jansen, M., Bali, M. & Akabas, M. H. Modular design of Cys-loop ligand-gated ion channels: functional 5-HT<sub>3</sub> and GABA<sub>A</sub> p1 receptors lacking the large cytoplasmic M3M4 loop. *J. Gen. Physiol.* **131**, 137–146 (2008).
- Hansen, S. B., Wang, H. L., Taylor, P. & Sine, S. M. An ion selectivity filter in the extracellular domain of Cys-loop receptors reveals determinants for ion conductance. *J. Biol. Chem.* **283**, 36066–36070 (2008).
- Sauguet, L. et al. Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO J.* **32**, 728–741 (2013).
- Gurba, K. N., Hernandez, C. C., Hu, N. & Macdonald, R. L. GABRB3 mutation, G32R, associated with childhood absence epilepsy alters α<sub>1</sub>β<sub>3</sub>γ<sub>2</sub>L γ-aminobutyric acid type A (GABA<sub>A</sub>) receptor expression and channel gating. *J. Biol. Chem.* **287**, 12083–12097 (2012).
- Sancar, F. & Czajkowski, C. A GABA<sub>A</sub> receptor mutation linked to human epilepsy (γ<sub>2</sub>R43Q) impairs cell surface expression of αβγ receptors. *J. Biol. Chem.* **279**, 47034–47039 (2004).
- Epi, K. C. et al. De novo mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).
- Bergmann, R., Kongsbak, K., Sorensen, P. L., Sander, T. & Balle, T. A Unified model of the GABA<sub>A</sub> receptor comprising agonist and benzodiazepine binding sites. *PLoS ONE* **8**, e52323 (2013).
- Hansen, S. B. et al. Structures of *Aplysia* AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24**, 3635–3646 (2005).
- Huang, S. et al. Complex between α-bungarotoxin and an α<sub>7</sub> nicotinic receptor ligand-binding domain chimera. *Biochem. J.* **454**, 303–310 (2013).
- Celie, P. H. et al. Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
- Newell, J. G., McDevitt, R. A. & Czajkowski, C. Mutation of glutamate 155 of the GABA<sub>A</sub> receptor β<sub>2</sub> subunit produces a spontaneously open channel: a trigger for channel activation. *J. Neurosci.* **24**, 11226–11235 (2004).
- Venkatachalan, S. P. & Czajkowski, C. A conserved salt bridge critical for GABA<sub>A</sub> receptor function and loop C dynamics. *Proc. Natl Acad. Sci. USA* **105**, 13604–13609 (2008).
- Wagner, D. A., Czajkowski, C. & Jones, M. V. An arginine involved in GABA binding and unbinding but not gating of the GABA<sub>A</sub> receptor. *J. Neurosci.* **24**, 2733–2741 (2004).
- Mukhtasimova, N., Free, C. & Sine, S. M. Initial coupling of binding to gating mediated by conserved residues in the muscle nicotinic receptor. *J. Gen. Physiol.* **126**, 23–39 (2005).
- Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nature Rev. Drug Discov.* **3**, 660–672 (2004).
- Lo, W. Y. et al. Glycosylation of β<sub>2</sub> subunits regulates GABA<sub>A</sub> receptor biogenesis and channel gating. *J. Biol. Chem.* **285**, 31348–31361 (2010).
- Miller, P. S., Da Silva, H. M. & Smart, T. G. Molecular basis for zinc potentiation at strychnine-sensitive glycine receptors. *J. Biol. Chem.* **280**, 37877–37884 (2005).
- Buhr, A. et al. Functional characterization of the new human GABA<sub>A</sub> receptor mutation β<sub>3</sub>(R192H). *Hum. Genet.* **111**, 154–160 (2002).
- Dellisanti, C. D., Yao, Y., Stroud, J. C., Wang, Z. Z. & Chen, L. Crystal structure of the extracellular domain of nAChR α1 bound to α-bungarotoxin at 1.94 Å resolution. *Nature Neurosci.* **10**, 953–962 (2007).
- Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).
- Akabas, M. H. Using molecular dynamics to elucidate the structural basis for function in pLGICs. *Proc. Natl Acad. Sci. USA* **110**, 16700–16701 (2013).
- Baulac, S. et al. First genetic evidence of GABA<sub>A</sub> receptor dysfunction in epilepsy: a mutation in the γ<sub>2</sub>-subunit gene. *Nature Genet.* **28**, 46–48 (2001).
- Lape, R., Plested, A. J., Moroni, M., Colquhoun, D. & Sivillotti, L. G. The α1K276E startle disease mutation reveals multiple intermediate states in the gating of glycine receptors. *J. Neurosci.* **32**, 1336–1352 (2012).
- Xu, M. & Akabas, M. H. Identification of channel-lining residues in the M2 membrane-spanning segment of the GABA<sub>A</sub> receptor α1 subunit. *J. Gen. Physiol.* **107**, 195–205 (1996).
- Cymes, G. D., Ni, Y. & Grosman, C. Probing ion-channel pores one proton at a time. *Nature* **438**, 975–980 (2005).
- Unwin, N. & Fujiyoshi, Y. Gating movement of acetylcholine receptor caught by plunge-freezing. *J. Mol. Biol.* **422**, 617–634 (2012).
- Horenstein, J. & Akabas, M. H. Location of a high affinity Zn<sup>2+</sup> binding site in the channel of α1β1 γ-aminobutyric acid<sub>A</sub> receptors. *Mol. Pharmacol.* **53**, 870–877 (1998).
- Rajendra, S. et al. Mutation of an arginine residue in the human glycine receptor transforms β-alanine and taurine from agonists into competitive antagonists. *Neuron* **14**, 169–175 (1995).
- Young, G. T., Zwart, R., Walker, A. S., Sher, E. & Millar, N. S. Potentiation of α<sub>7</sub> nicotinic acetylcholine receptors via an allosteric transmembrane site. *Proc. Natl Acad. Sci. USA* **105**, 14686–14691 (2008).
- daCosta, C. J., Free, C. R., Corradi, J., Bouzat, C. & Sine, S. M. Single-channel and structural foundations of neuronal α<sub>7</sub> acetylcholine receptor potentiation. *J. Neurosci.* **31**, 13870–13879 (2011).
- Shan, Q., Hadrill, J. L. & Lynch, J. W. Ivermectin, an unconventional agonist of the glycine receptor chloride channel. *J. Biol. Chem.* **276**, 12556–12564 (2001).
- Jensen, M. L. et al. The β subunit determines the ion selectivity of the GABA<sub>A</sub> receptor. *J. Biol. Chem.* **277**, 41438–41447 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Malinauskas, Y. Kong and staff at Diamond Light Source beamlines I03 and I24 for synchrotron assistance; K. Harlos and T. Walter for technical support with crystallization; G. Schertler and J. Standfuss for advice concerning the Rho-1D4 affinity purification method; T. Nakagawa for electron microscopy sample examination; F. Ashcroft, S. Tucker, M. Clausen and P. Proks for access to electrophysiology equipment and assistance with electrophysiological recordings; J. McIlhinney, M. Sansom, L. Carpenter, S. Newstead, I. de Moraes and members of the Aricescu laboratory for discussions; E.Y. Jones, D.J. Stuart and C. Siebold for reading the manuscript. This work was supported by grants from the Wellcome Trust (OXION: Ion channels and Disease Initiative, 084655), the UK Medical Research Council (MRC) G0700232 and the Royal Society (RG090810). Further support from the Wellcome Trust Core Award Grant Number 090532/Z/09/Z is acknowledged. P.S.M. was a Wellcome Trust OXION Training Fellow. A.R.A. is an MRC Senior Research Fellow.

**Author Contributions** The authors have jointly contributed to the project design, data analysis and manuscript preparation. Experimental work was performed by P.S.M. (protein expression, purification, crystallization, ligand binding assays and electrophysiology) and A.R.A. (crystallography).

**Author Information** The coordinates and the structure factors have been deposited in the Protein Data Bank under the accession code 4COF. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details can be found in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R.A. ([radu@strubi.ox.ac.uk](mailto:radu@strubi.ox.ac.uk)) or P.S.M. ([paul@strubi.ox.ac.uk](mailto:paul@strubi.ox.ac.uk)).

# X-ray structure of the mouse serotonin 5-HT<sub>3</sub> receptor

Ghérici Hassaine<sup>1\*†</sup>, Cédric Deluz<sup>1\*</sup>, Luigino Grasso<sup>1</sup>, Romain Wyss<sup>1</sup>, Menno B. Tol<sup>1</sup>, Ruud Hovius<sup>1</sup>, Alexandra Graff<sup>2</sup>, Henning Stahlberg<sup>2</sup>, Takashi Tomizaki<sup>3</sup>, Aline Desmyter<sup>4</sup>, Christophe Moreau<sup>5,6,7</sup>, Xiao-Dan Li<sup>8</sup>, Frédéric Poitevin<sup>9</sup>, Horst Vogel<sup>1</sup> & Hugues Nury<sup>1,5,6,7</sup>

**Neurotransmitter-gated ion channels of the Cys-loop receptor family mediate fast neurotransmission throughout the nervous system. The molecular processes of neurotransmitter binding, subsequent opening of the ion channel and ion permeation remain poorly understood. Here we present the X-ray structure of a mammalian Cys-loop receptor, the mouse serotonin 5-HT<sub>3</sub> receptor, at 3.5 Å resolution. The structure of the proteolysed receptor, made up of two fragments and comprising part of the intracellular domain, was determined in complex with stabilizing nanobodies. The extracellular domain reveals the detailed anatomy of the neurotransmitter binding site capped by a nanobody. The membrane domain delimits an aqueous pore with a 4.6 Å constriction. In the intracellular domain, a bundle of five intracellular helices creates a closed vestibule where lateral portals are obstructed by loops. This 5-HT<sub>3</sub> receptor structure, revealing part of the intracellular domain, expands the structural basis for understanding the operating mechanism of mammalian Cys-loop receptors.**

The complex mental capacities and motor behaviours of mammals rely on the evolution of distinguished central and peripheral nervous systems where ligand-gated ion channels facilitate fast intercellular signalling. The serotonin-gated 5-HT<sub>3</sub> receptor<sup>1,2</sup>, which belongs to a family of pentameric neurotransmitter-gated ion channels termed Cys-loop receptors<sup>3,4</sup>, is the target of potent drugs that alleviate chemotherapy-induced and post-operative nausea and vomiting, and may enable personalized therapies for irritable bowel syndrome and various psychiatric disorders<sup>5</sup>. To date, five 5-HT<sub>3</sub> receptor subunits, named A to E, have been identified in humans while rodents only possess the A and B subunits. Heterologous expression allowed the characterization of the homopentameric A and heteropentameric AB receptors; the role of subunits C to E remains to be elucidated<sup>2,5</sup>. A and B subunits are expressed widely throughout the brain, for example in the hippocampus and amygdala, and in peripheral organs like the gastrointestinal tract, but the exact subunit composition of native human receptors remains elusive. The Cys-loop receptor family also encompasses the nicotinic acetylcholine (nACh) receptors, which together with 5-HT<sub>3</sub> receptors are cation-selective and promote excitatory signals; it also includes the  $\gamma$ -aminobutyric acid (GABA) and glycine receptors, which are inhibitory anion-selective channels. Cys-loop receptors have a major role in virtually all brain functions, are involved in numerous pathologies and are thus targets of many psychoactive and therapeutic compounds (including nicotine, alcohol, benzodiazepines, barbiturates, steroids and local and general anaesthetics).

These receptors function as allosteric signal transducers<sup>6</sup> across the plasma membrane: upon binding of one or more neurotransmitter molecules to an extracellular site, the receptors undergo complex conformational transitions that result in transient opening of an intrinsic transmembrane channel<sup>4,7–9</sup>. Yet, structural and mechanistic details of these processes remain speculative as high-resolution structures of mammalian

receptors are missing, due to difficulties in heterologous expression, purification and crystallization of these proteins.

Present understanding of Cys-loop receptor structure is based on electron microscopy imaging of *Torpedo marmorata* nACh receptors (*Torpedo* nAChR)<sup>10,11</sup>, on X-ray structures of the bacterial homologues GLIC<sup>12–16</sup> and ELIC<sup>17,18</sup>, and of the *Caenorhabditis elegans* homologue GluCl<sup>19</sup>. A prototypical receptor contains three distinct domains: a large,  $\beta$ -sheet-rich, extracellular domain comprising the ligand-binding site(s), an  $\alpha$ -helical pore-forming domain, and an intracellular domain interacting with cellular scaffolding proteins. The intracellular domain, specific to metazoa, is absent in the structures of GLIC, ELIC and GluCl. To obtain a high-resolution structure of a mammalian neurotransmitter-gated ion channel comprising all three domains, we crystallized the mouse homopentameric 5-HT<sub>3A</sub> receptor.

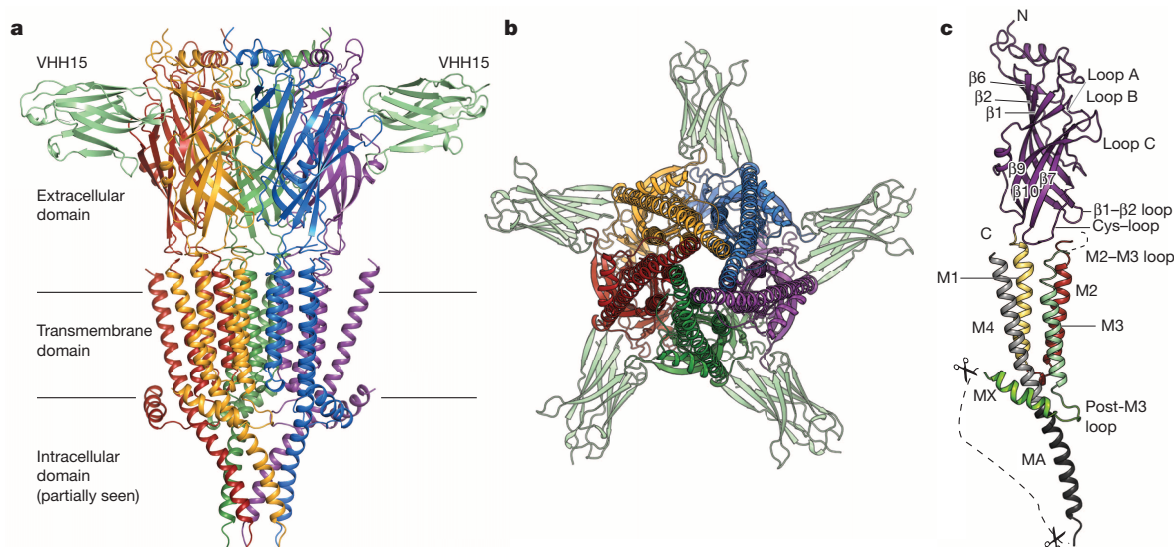
## Crystallization and architecture

The 5-HT<sub>3</sub> receptor was expressed in human embryonic kidney 293 (HEK 293) cells<sup>20</sup>. First screens for the full-length receptor yielded crystals diffracting X-rays to 20 Å. We carried out limited proteolysis experiments. Trypsin cleaved ~50 residues, yielding a receptor composed of two fragments per subunit (Fig. 1c and Extended Data Fig. 1). This split 5-HT<sub>3</sub> receptor remained pentameric with an unchanged ligand-binding profile (Extended Data Fig. 1a–c, e). After reconstitution in liposomes, it showed serotonin-activated single-channel activity (Extended Data Fig. 1d), consistent with reports that split glycine and 5-HT<sub>3</sub> receptors are functional<sup>21</sup>. The currents attributed to split receptors presented a single channel conductance of ~30 pS, substantially higher than the <1 pS wild-type receptor conductance, indicating that alteration of the intracellular domain impacts permeation properties. The diffraction of split 5-HT<sub>3</sub> receptor crystals plateaued at 7 Å resolution. We used

<sup>1</sup>Laboratory of Physical Chemistry of Polymers and Membranes, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. <sup>2</sup>Center for Cellular Imaging and NanoAnalytics, Biozentrum, University of Basel, CH-4058 Basel, Switzerland. <sup>3</sup>Swiss Light Source, Paul Scherrer Institute, CH-5234 Villigen, Switzerland. <sup>4</sup>Architecture et Fonction des Macromolécules Biologiques, CNRS UMR 7257 and Université Aix-Marseille, F-13288 Marseille, France. <sup>5</sup>Université Grenoble Alpes, IBS, F-38000 Grenoble, France. <sup>6</sup>CNRS, IBS, F-38000 Grenoble, France. <sup>7</sup>CEA, DSV, IBS, F-38000 Grenoble, France. <sup>8</sup>Laboratory of Biomolecular Research, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. <sup>9</sup>Unité de Dynamique Structurale des Macromolécules, Institut Pasteur, CNRS UMR3528, F-75015 Paris, France. <sup>†</sup>Present address: Theranox, 163 Avenue de Luminy, 13288 Marseille, France.

\*These authors contributed equally to this work.





**Figure 1 | Architecture of the 5-HT<sub>3</sub> receptor in complex with VHH15.** **a**, Cartoon representation viewed parallel to the plane of the membrane. Only two out of five VHH15 molecules are shown (pale green). **b**, Intracellular view, perpendicular to the membrane, of the complex. **c**, A single subunit of the 5-HT<sub>3</sub> receptor is viewed parallel to the plane of the membrane. The large

fragment (extracellular domain, first three transmembrane helices, post-M3 and MX) is in colour, the small fragment (helices MA and M4) is in greyscale. The segment removed by proteolysis is indicated as a dashed line. Secondary structure elements and important loops are noted.

llama-derived single-chain antibodies<sup>22</sup> (termed nanobody or VHH) as crystallization chaperones. VHH15 formed a stable complex of sub-nanomolar affinity with the 5-HT<sub>3</sub> receptor; crystals of the complex diffracted X-rays to 3.5 Å resolution. The structural model, comprising residues 9–275, 280–334 and 399–459 of the receptor, has good quality electron density maps for most regions (Extended Data Figs 2a and 3).

The 5-HT<sub>3</sub> receptor is bullet-shaped, with subunits arranged in five-fold symmetry around a central ion pathway perpendicular to the membrane plane (see Figs 1, 2 and Extended Data Fig. 2c–e for topology and nomenclature). The cylindrical extracellular and membrane domains have similar architecture to ELIC, GLIC and GluCl. Five VHH15 molecules bind radially to the interfaces between subunits, like fins, at the level of the serotonin binding sites. The conical intracellular domain comprises an additional portion proximal to the membrane, the post-M3 loops and the MX helices. The intracellular MA helices are twisted, forming a tight bundle at the receptor's tip.

## Ion permeation pathway

Extracellular ions entering the 60-Å high and 20-Å wide extracellular vestibule (Fig. 2) encounter a first constriction at the level of residues D105 and K108 of loop β4–β5, which is narrower than those found in homologous structures (Extended Data Fig. 4a–c, see also multiple sequence alignment in the Supplementary Information). Removing the negative charge at position 105 markedly reduced channel conductance<sup>23</sup>, concordant with reports that charged residues in the vestibule influenced the elementary conductance of α1-nACh<sup>24</sup> or glycine and GABA<sub>A</sub><sup>25</sup> receptors. The side chains of K108 define the constriction because they adopt, in the crystal structure, an extended conformation to coordinate a small molecule from the crystallization liquor, best modelled as a sulphate ion. Other charged residues lining the vestibule include R14, D17, K24, K25, D91, D52, D53, K54 and E186 (Fig. 2a). Below the constriction, the vestibule exhibits an electronegative surface capable of increasing the local concentration of cations (Fig. 2c and Extended Data Fig. 4d).

Ions then enter the ~40-Å-long transmembrane pore itself. The M2 helices line the pore, following the common organization of the family<sup>3</sup>. Side chains of identical residues from each subunit form rings building the pore lumen surface at positions –1', 2', 6', 9', 13', 16', 17' and 20' (Fig. 3a). Substituted cysteine accessibility mutagenesis identified these residues to be accessible from the bulk water phase as well as additional residues at positions 7', 10', 12' and 15' that do not face the pore in the

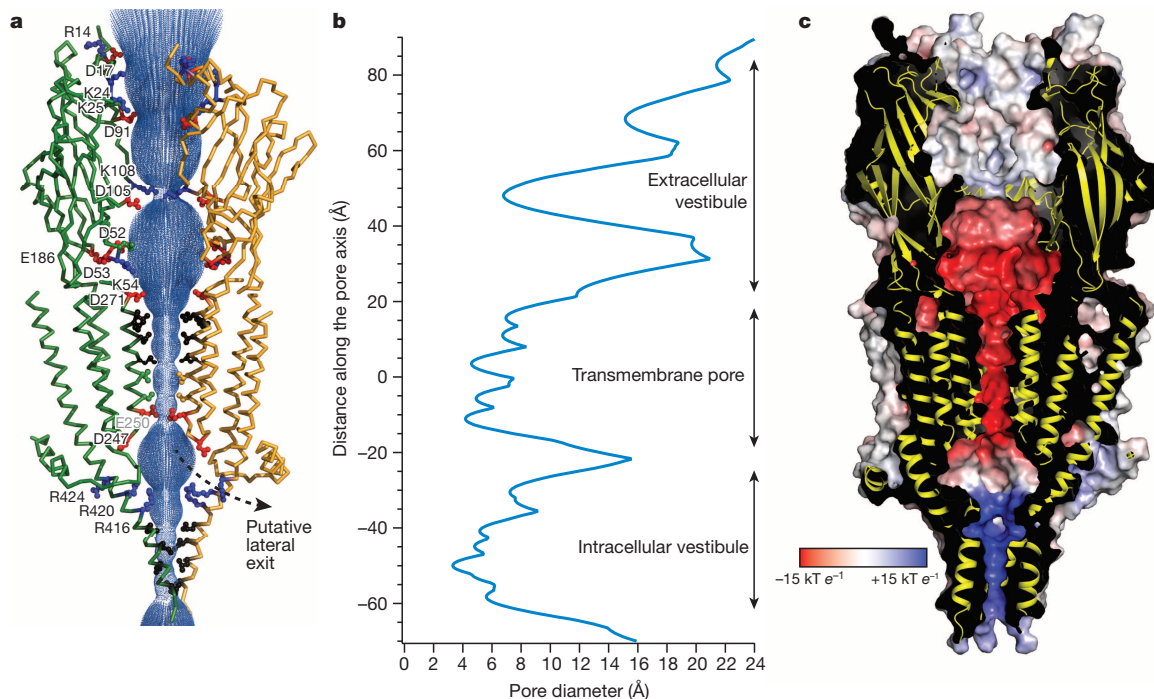
structure<sup>26,27</sup>. Conformational flexibility of the top of the M2 helix during gating, structurally documented for GLIC<sup>14,16</sup>, could explain accessibility at positions 10', 12' and 15'.

In the extracellular half of M2, rings of hydrophobic residues at positions 16', 13' and 9' define a funnel with a 4.6-Å hydrophobic constriction at the level of L9' (L260). In this region, believed to contain the main channel gate, the 5-HT<sub>3</sub> receptor pore state is not clearly defined. On the one hand, the width of the pore is smaller than that of homologous X-ray structures deemed open (GLIC<sub>o</sub><sup>13</sup> and GluCl<sup>19</sup>) (Fig. 3b, c). The constriction might form a hydrophobic barrier<sup>28</sup> too narrow for cations to pass while retaining their full hydration shell, which could suggest the pore to be in a non-conducting state. On the other hand, the global backbone conformation and the pore radius at the backbone level resemble those of open channels (GLIC<sub>o</sub> and GluCl) but clearly differ from those of closed channels (ELIC<sup>17</sup>, resting/closed GLIC<sub>c</sub><sup>16</sup> and locally closed GLIC<sub>LC</sub><sup>14</sup>). In addition, slight fluctuations in the local pore conformation and hydration might allow ion passage<sup>15,29</sup>. The *Torpedo* nAChR pore<sup>10,11</sup> is off-register by one helix turn<sup>13,19</sup>, precluding detailed comparison.

Continuing along the ion pathway, two rings of polar residues at positions 2' and 6' are encountered with a constriction of 5 Å at the level of S2'. Increased electron density is present in the pore between S2' and T6', which might arise from the presence of a cation and is located close to ion sites observed in GLIC<sup>15</sup> and GluCl<sup>19</sup>. Both ends of the transmembrane pore are lined with rings of negative charges: D20' at the top, D–4' and E–1' at the bottom, the latter being a major determinant of ionic selectivity<sup>30,31</sup>. The side chain of E–1' is not seen in the density maps, possibly because it fluctuates between alternate conformations<sup>15,32</sup>, and was modelled as the most plausible rotamer.

## The intracellular domain

After passing the transmembrane pore, the cations access the intracellular domain. Vertebrate receptors possess an intracellular domain of 70 to 150 residues important for their trafficking and clustering at the synapse<sup>3,33</sup> and for gating kinetics<sup>34</sup>. In 5-HT<sub>3</sub> receptors, this domain is a major determinant of channel conductance<sup>35,36</sup>. Structural information has remained limited: the *Torpedo* nAChR structure shows an intracellular MA helix but the intracellular domain is absent in GLIC and ELIC; in GluCl it was replaced by a short loop connecting M3 and M4.



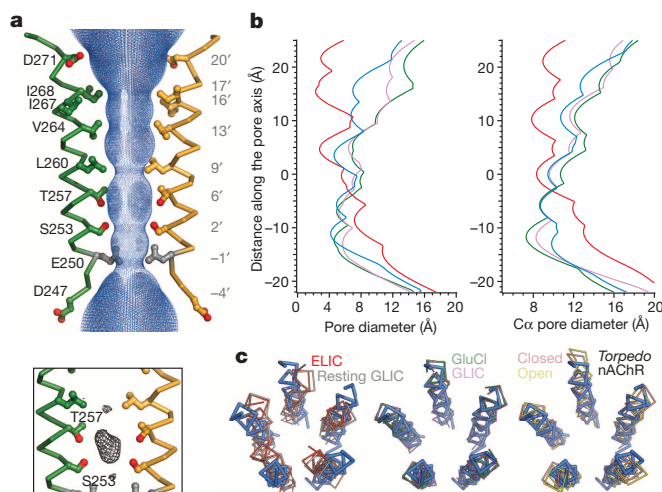
**Figure 2 | The ion permeation pathway.** **a**, Global view of the ion permeation pathway. Two subunits are shown as ribbons with main residues lining the permeation pathway as sticks (negative in red, positive in blue, hydrophobic in black). The solvent-accessible surface is represented as blue dots. **b**, Pore profile

The 5-HT<sub>3</sub> receptor structure depicts the first 20 residues of this domain, which form a loop, and the short MX helix clamping MA–M4 (Figs 1, 4 and Extended Data Fig. 5). For MX residues 317 to 334 we could trace the main polypeptide chain, but side-chain densities were not sufficiently resolved for model building (Extended Data Fig. 3e)

for the whole receptor (transmembrane pore close-up in Fig. 3). **c**, Solvent-accessible electrostatic potential mapped on the receptor surface, colour-coded from  $-15 \text{ kT e}^{-1}$  to  $15 \text{ kT e}^{-1}$  (red to blue).

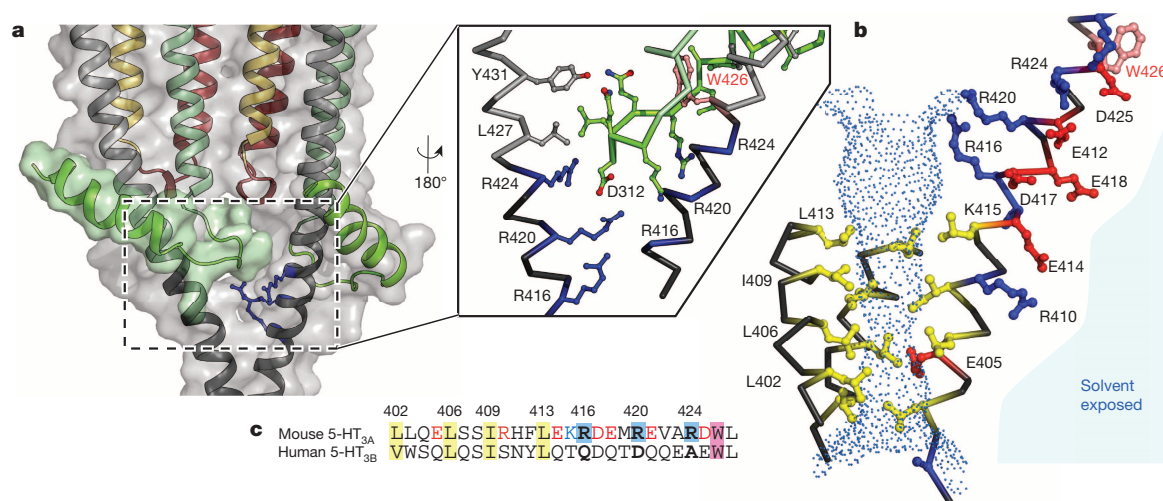
indicating high intrinsic flexibility. The structure also shows the last 25 residues forming the MA helix as a continuous extension of M4. Notably, a stretch of 30 residues of MA and M4 (homologous to residues 410–440) was found to be responsible for different gating kinetics of nACh receptors containing either  $\gamma$  or  $\epsilon$  subunits<sup>34</sup>, underlining the contribution of that region to gating properties. Missing from the structure are 62 residues, corresponding mostly to the trypsin-digested stretch. The end of MX and the beginning of MA are separated by 61 Å. In the Cys-loop receptor family this missing part contains interaction sites for proteins modulating activity, assembly, trafficking and clustering<sup>33</sup>.

The position of the post-M3 loop relative to MA–M4 is restrained by the residue W426, conserved among cationic receptors. Globally the intracellular domain creates a vestibule exposing on its inner surface residues of the post-M3 loop (H309, K310 and D312), of the M1–M2 loop D247 (D–4') and of the MA helix (R416, R420) (Extended Data Fig. 5c). Lateral portals between MA helices, seen in the *Torpedo* nAChR structure<sup>10</sup>, are here blocked by the post-M3 loop leaving only a narrow tunnel, at its thinnest just 3.3 Å in diameter (Fig. 4a and Extended Data Fig. 5a, b, e). Further down from the membrane, the MA helical bundle tightens and rings of hydrophobic residues L402, L406, I409 and L413 stack up to create a 17-Å-long narrow zone, with a minimum diameter of 4.2 Å (Figs 2b and 4b). The structure described here therefore does not offer an exit pathway for the ions. Conformational changes have to take place for the ions to exit, through an opening believed to be larger than the transmembrane pore selectivity filter<sup>37</sup>, either at the lateral portals or along the channel axis. The tight hydrophobic juncture at the bottom of the MA helices, together with mutagenesis data that identify the upper rather than the lower MA helix as participating to the ion pathway<sup>38</sup>, favour the portals scenario. However, whether the conformation of the post-M3 loop and of the MX helix in the detergent-solubilized, trypsin-treated, crystallized receptors do accurately represent a physiological conformation of intact receptors at the plasma membrane remains to be investigated. The degree of intrinsic flexibility of the post-M3 loop and MX helix, whether they move, and/or whether the portals widen to open the way for ions—for example as a consequence of a global twist motion during gating—are also unresolved questions.



**Figure 3 | Pore conformation.** **a**, View in the membrane plane of the ion pore of two M2 helices. The pore lumen is represented by blue dots and exposed side chains are depicted as sticks. The side chain of E250, not seen in electron density maps, is coloured grey. Red dots are oxygen atoms. Inset: central density in the  $F_o - F_c$  map contoured at  $3\sigma$  is represented as black mesh. **b**, All-atom and  $C\alpha$  pore profiles for the 5-HT<sub>3</sub> receptor (blue), ELIC (red), GLIC (magenta) and GluCl (green) along their common axis. The common origin is the  $C\alpha$  of residue 6'. **c**, Superposition of the M2 helix bundle (in blue) with the closed ELIC (2.85) and resting GLIC (2.72), the open GluCl (0.80) and GLIC (0.86) and the closed (1.27) and open (1.44) *Torpedo* nAChR. Values in parentheses are root-mean-square deviations (r.m.s.d.) in Å at the backbone level.





**Figure 4 | The intracellular domain.** **a**, Side view of two subunits, at the transmembrane and intracellular domain level, as cartoon and surface representation. The post-M3 loop and MX helix of one subunit is highlighted with green surface to show how it obstructs the portal between the MA helices. The three conductance-defining arginine residues are depicted as sticks. Inset: enlarged view from the intracellular vestibule showing two MA–M4s (grey backbone) and one post-M3 loop (green backbone) interactions. Red dots,

oxygen atoms; blue dots, nitrogen atoms. **b**, View of the intracellular tight hydrophobic constriction (yellow sticks, pore surface as blue dots) and of the striking charge pattern of the MA helix (one subunit only is represented with red and blue sticks for negatively and positively charged residues, respectively). **c**, Sequence alignment of the mouse 5-HT<sub>3A</sub> and human 5-HT<sub>3B</sub> sequences. The arginine triplet (bold) is present in A and absent in B subunits. Colours correspond to the structural snapshots.

From the observation that conductances of 5-HT<sub>3A</sub> homomeric receptors and 5-HT<sub>3AB</sub> heteromeric receptors were markedly different (0.4 pS versus 16 pS, respectively), a seminal study identified a triplet of arginine residues within helix MA of the A subunit responsible for the low conductance of homomeric receptors<sup>35</sup> (Fig. 4c). Combined mutation of 5-HT<sub>3A</sub> receptor residues R416, R420 and R424 to their counterparts glutamine, aspartic acid and alanine of the 5-HT<sub>3B</sub> receptor yielded a conductance jump to the 20–40 pS range. R416 faces the pore lumen; R420, within salt-bridge distance to D312, can point into or out of the vestibule; and R426 is exterior to the post-M3 loop and more oriented towards the cytosol (the arginine side chains are not all well resolved, but their  $\alpha$ -carbon ( $C\alpha$ ) positions are; Extended Data Figs 2 and 5). The three conductance-defining arginines are surrounded by negatively charged residues (E414, D417, E418, E421 and D425) that create a ladder of negative charges along the outer face of the MA helix (Fig. 4d). Removal of the negative charges at positions 414, 417 and to a lesser extent at positions 421 and 425 increased the conductance<sup>39</sup>. Homology modelling based on the *Torpedo* nAChR structure suggested that the arginine residues could either line the portals and create an electrostatic energy barrier<sup>35,36,38</sup>, or they could engage in inter-subunit salt bridges with neighbouring negatively charged residues, conferring structural rigidity that is adverse for translocation<sup>39</sup>. Comparing our resolved intracellular domain to that of the *Torpedo* nAChR structure, we found shifts in register between the sequence alignment and the structural superimposition (Extended Data Fig. 6a–d) and distinct organizations of the MA helix bundle (Extended Data Fig. 6e, f). Our structure thus provides a new framework to interpret the functional data.

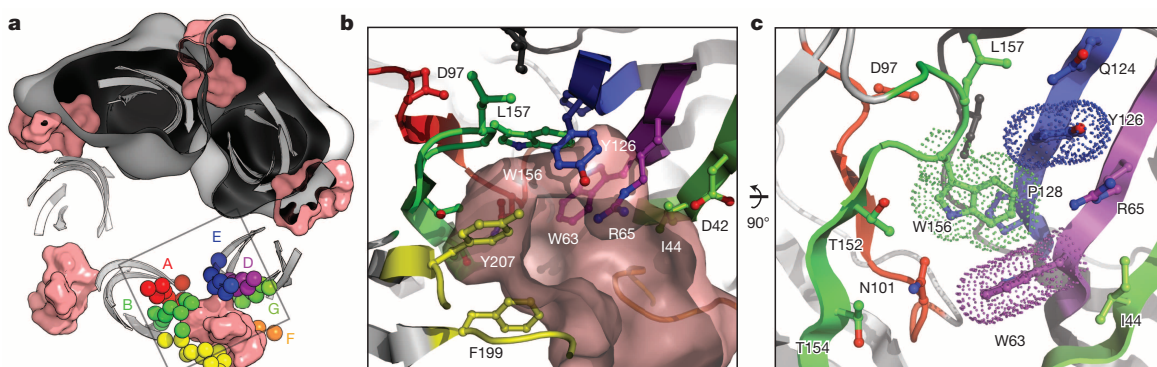
## Neurotransmitter binding site

The binding site anatomy has been explored by a wealth of functional and biochemical studies on the 5-HT<sub>3</sub> and other Cys-loop receptors<sup>2,3</sup>, and by structural studies of acetylcholine-binding proteins<sup>40–43</sup> (AChBP) serving as surrogates for the extracellular domain of Cys-loop receptors. Each of the five equivalent serotonin binding sites of the 5-HT<sub>3</sub> receptor are located between two subunits in an electronegative cleft (Fig. 5a and Extended Data Fig. 7). Loops A, B and C from the principal subunit and portions of  $\beta$ -strands D, E, G and loop F from the complementary subunit contribute to the site. A, B, D and E elements are more internal and connected through a network of conserved residues (sometimes referred

to as a ‘stabilizing box’<sup>44</sup>; Extended Data Fig. 10), while loops C and F are peripheral and less restrained.

The conformation of loop C was postulated to reflect the functional state of the receptor, contracted in agonist-bound AChBP structures and open in antagonist-bound structures<sup>41,43</sup>. Here loop C adopts an intermediate conformation, wrapped by VHH15 loops CDR1, CDR2 and D–E, while CDR3 interacts mainly with loop  $\beta$ 8– $\beta$ 9 of the complementary subunit (Extended Data Fig. 7g, h). Each VHH15 buries a surface of 583 and 431 Å<sup>2</sup> on the principal and complementary subunit, respectively. Of note, VHH15 antibodies contribute to all packing contacts. Owing to its contact to the binding site, we investigated whether VHH15 did have an influence on the function of the receptor. Electrophysiology measurements revealed VHH15 as a potent inhibitor of serotonin-elicited ion currents (Extended Data Fig. 1f–i): dose-response measurements performed on *Xenopus* oocytes yielded an IC<sub>50</sub> (dose concentration at half-maximal inhibition) of 29 nM, with similar effects in HEK 293 cells. Among all screened nanobodies, VHH15 was one of the most potent inhibitors. In radioligand binding competition assays, nanomolar concentrations of VHH15 inhibited half of the [<sup>3</sup>H]GR65630 antagonist binding. A sub-nanomolar dissociation constant ( $K_d$ ) was derived from surface plasmon resonance binding experiments (Extended Data Fig. 1j, k).

The remainder of the binding site appears unaffected by VHH15 and displays an archetypal organization (Fig. 5b, c and Extended Data Fig. 7). Aromatic residues from loop B (W156), loop D (W63) and loop E (Y126), together with F199 and Y207 from loop C, create a ~10-Å-sized aromatic ‘box’. At the centre of the box, increased electron density is present in  $F_o - F_c$  maps at  $\sim 4\sigma$ , suggesting the binding of a small molecule from the crystallization liquor. W156, which was shown to have cation– $\pi$  interaction with serotonin<sup>44</sup>, forms the back wall of the box. This cation– $\pi$  interaction, initially postulated for acetylcholine in nACh receptors<sup>45</sup>, is conserved within the family. W156 protrudes from the principal subunit at the tip of loop B and interacts with three different  $\beta$ -strands of the complementary subunit ( $\beta$ 2,  $\beta$ 5 and  $\beta$ 6). Loop B, a component of the stabilizing box, is a rigid structure, part of which is conferred by L157 and H158 interactions with D97, as previously evidenced by non-natural amino acid mutagenesis<sup>46</sup>. Because of this rigidity, W156 may act as a ball-joint during the quaternary reorganization involved in gating. R65 in loop D, which can form a cation– $\pi$



**Figure 5 | Neurotransmitter binding site.** **a**, Extracellular view of the binding sites at the interface between subunits. Binding cleft surfaces are represented in pink, the extracellular domain core atoms are depicted in cartoon representation, with smooth surfaces for two of them. For one of the 5 binding sites, the binding loops A–G are indicated by coloured spheres for  $\alpha$  atoms. **b**, Enlarged extracellular view with residues important for agonist binding

interaction with the indazole moiety of granisetron<sup>42</sup>, presents here an extended conformation. It corresponds to R56 in GluCl that interacts with glutamate. Other internal residues pointing towards the site include N101 (loop A), T152 and T154 (loop B), while peripheral ones include D42, I44 (loop G), D177, S179, I180 (loop F) and residues from loop C.

### Interface and quaternary structure

At the interface between the extracellular and the transmembrane domains, the crucial arginine R218 (refs 47, 48) of the pre-M1 loop is surrounded by three negatively charged residues E53 (loop  $\beta$ 1– $\beta$ 2, highly conserved), D145 (Cys-loop, strictly conserved) and E186 (loop  $\beta$ 8– $\beta$ 9, well conserved in the cationic branch of the family, in the so-called GEW motif). These charged residues are sandwiched between the conserved aromatic residue W187 of loop  $\beta$ 8– $\beta$ 9 above and the 142-FPF-144 motif of the Cys-loop below (Extended Data Fig. 8). Notably, residues equivalent to E53, E186, W187 and R218 were shown to inter-dependently contribute to signal transduction<sup>49</sup>. The sandwich arrangement seems to be conserved among Cys-loop receptor structures and its location might be an indicator of the state of the channel (Extended Data Fig. 8). The Cys-loop penetrates into the four-helix bundle of the transmembrane domain, and is located at the same height as the M2–M3 loop. The tip of the M2–M3 loop is disordered and residues 276–279 do not appear in electron density maps, indicating that the conformation captured is compatible with a degree of flexibility in this loop. The helical structure of M1 is broken at the conserved proline P230 (Extended Data Fig. 8d), close to the interface, providing flexibility to the extracellular domain orientation relative to the transmembrane domain. The relative orientation of domains is unique for each receptor, as illustrated by the substantial spread of transmembrane domains when superposing on a subunit extracellular domain (Extended Data Fig. 9a–c). Finally, when ELIC, GLIC, GluCl and the 5-HT<sub>3</sub> receptor structures are superposed, the latter appears as if it continued a global motion going from the closed to the apparently open structures (Extended Data Fig. 9d–g). This impression is confirmed when looking at a superposition of conserved motifs only, minimizing bias when comparing different species (Extended Data Fig. 10).

### Conclusion and future directions

The structure of the split 5-HT<sub>3</sub> receptor delivers a number of interesting, unexpected features. First, it crystallizes in complex with VHH15, a potent inhibitor, which is therefore expected to stabilize a non-conducting channel conformation. Second, although the transmembrane pore bears a minimal diameter intermediate between the apparently open and the tightly closed channels, the backbone conformation of the M2 helix

represented as sticks, and the binding cleft surface depicted in pink. The image corresponds to the rectangle in panel **a**. **c**, Enlarged view parallel to the membrane plane. Van der Waals surface is depicted as dots for aromatic box residues and loop C is removed for clarity. **b**, **c**, Red dots, oxygen atoms; blue dots, nitrogen atoms.

bundle resembles that of GLIC and GluCl and differs distinctly from that of resting/closed GLIC and closed ELIC. Third, the split receptor displays a closed intracellular domain tightly coupled to the membrane domain via the continuous MA–M4 helices and via the post-M3 loop obstructing the lateral portals. We anticipate our structure to be the starting point for complementary research exploring its dynamics and unambiguously assigning the crystal structure to a physiological conformation at the membrane.

Cys-loop receptors are major targets for therapeutics to treat neurological disorders, and the 5-HT<sub>3</sub> receptor structure opens new avenues for rational drug design. Furthermore, as nanobodies may cross the blood–brain barrier<sup>50</sup>, they are potential diagnostic and therapeutic tools. Our finding that a particular nanobody binds at the subunits' extracellular interface opens the possibility to specifically target Cys-loop receptor subtypes, which is a crucial challenge in developing new active compounds.

### METHODS SUMMARY

The 5-HT<sub>3</sub> receptor was expressed in inducible HEK 293 cell lines, solubilized in C<sub>12</sub>E<sub>9</sub> detergent and purified by Streptactin affinity chromatography. After deglycosylation and limited trypsin proteolysis, it was submitted to size-exclusion chromatography and concentrated to ~4 mg ml<sup>−1</sup>. A mixture containing a 1:2 ratio of receptor and VHH15 (~4 mg ml<sup>−1</sup>) supplemented with 0.56 mM of Cymal-6 was used for vapour diffusion crystallization with a reservoir solution containing 20–25% PEG 10000, 100 mM Na<sub>2</sub>SO<sub>4</sub> and 100 mM Tris pH 8.5. Phases were obtained by molecular replacement. Biophysical and functional characterizations were performed using surface plasmon resonance, voltage-clamp electrophysiology of HEK 293 cells or *Xenopus* oocytes, and radioligand binding assays.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 January; accepted 2 June 2014.

Published online 3 August 2014.

1. Maricq, A. V., Peterson, A. S., Brake, A. J., Myers, R. M. & Julius, D. Primary structure and functional expression of the 5HT<sub>3</sub> receptor, a serotonin-gated ion channel. *Science* **254**, 432–437 (1991).
2. Lummis, S. C. R. 5-HT<sub>3</sub> receptors. *J. Biol. Chem.* **287**, 40239–40245 (2012).
3. Thompson, A. J., Lester, H. A. & Lummis, S. C. R. The structural basis of function in Cys-loop receptors. *Q. Rev. Biophys.* **43**, 449–499 (2010).
4. Corringer, P.-J. *et al.* Structure and pharmacology of pentameric receptor channels: from bacteria to brain. *Structure* **20**, 941–956 (2012).
5. Walstab, J., Rappold, G. & Niesler, B. 5-HT<sub>3</sub> receptors: role in disease and target of drugs. *Pharmacol. Ther.* **128**, 146–169 (2010).
6. Changeux, J.-P. 50 years of allosteric interactions: the twists and turns of the models. *Nature Rev. Mol. Cell Biol.* **14**, 819–829 (2013).
7. Schmauder, R., Kosanic, D., Hovius, R. & Vogel, H. Correlated optical and electrical single-molecule measurements reveal conformational diffusion from ligand



- binding to channel gating in the nicotinic acetylcholine receptor. *ChemBioChem* **12**, 2431–2434 (2011).
8. daCosta, C. J. B. & Baenziger, J. E. Gating of pentameric ligand-gated ion channels: structural insights and ambiguities. *Structure* **21**, 1271–1283 (2013).
  9. Auerbach, A. The gating isomerization of neuromuscular acetylcholine receptors. *J. Physiol. (Lond.)* **588**, 573–586 (2010).
  10. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
  11. Unwin, N. & Fujiyoshi, Y. Gating movement of acetylcholine receptor caught by plunge-freezing. *J. Mol. Biol.* **422**, 617–634 (2012).
  12. Hilf, R. J. C. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457**, 115–118 (2009).
  13. Bocquet, N. *et al.* X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* **457**, 111–114 (2009).
  14. Prevost, M. S. *et al.* A locally closed conformation of a bacterial pentameric proton-gated ion channel. *Nature Struct. Mol. Biol.* **19**, 642–649 (2012).
  15. Sauguet, L. *et al.* Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO J.* **32**, 728–741 (2013).
  16. Sauguet, L. *et al.* Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation. *Proc. Natl Acad. Sci. USA* **111**, 966–971 (2014).
  17. Hilf, R. J. C. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
  18. Spurny, R. *et al.* Pentameric ligand-gated ion channel ELIC is activated by GABA and modulated by benzodiazepines. *Proc. Natl Acad. Sci. USA* **109**, E3028–E3034 (2012).
  19. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* **474**, 54–60 (2011).
  20. Hassaine, G. *et al.* Large scale expression and purification of the mouse 5-HT<sub>3</sub> receptor. *Biochim. Biophys. Acta* **1828**, 2544–2552 (2013).
  21. Haeger, S. *et al.* An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nature Struct. Mol. Biol.* **17**, 90–98 (2010).
  22. Muyldermans, S. Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).
  23. Livesey, M. R., Cooper, M. A., Lambert, J. J. & Peters, J. A. Rings of charge within the extracellular vestibule influence ion permeation of the 5-HT<sub>3A</sub> receptor. *J. Biol. Chem.* **286**, 16008–16017 (2011).
  24. Hansen, S. B., Wang, H. L., Taylor, P. & Sine, S. M. An ion selectivity filter in the extracellular domain of Cys-loop receptors reveals determinants for ion conductance. *J. Biol. Chem.* **283**, 36066–36070 (2008).
  25. Moroni, M., Meyer, J. O., Lahmann, C. & Sivilotti, L. G. In glycine and GABA<sub>A</sub> channels, different subunits contribute asymmetrically to channel conductance via residues in the extracellular domain. *J. Biol. Chem.* **286**, 13414–13422 (2011).
  26. Reeves, D. C., Goren, E. N., Akabas, M. H. & Lummis, S. C. Structural and electrostatic properties of the 5-HT<sub>3</sub> receptor pore revealed by substituted cysteine accessibility mutagenesis. *J. Biol. Chem.* **276**, 42035–42042 (2001).
  27. Panicker, S., Cruz, H., Arrabit, C. & Slesinger, P. A. Evidence for a centrally located gate in the pore of a serotonin-gated ion channel. *J. Neurosci.* **22**, 1629–1639 (2002).
  28. Beckstein, O. & Sansom, M. S. P. A hydrophobic gate in an ion channel: the closed state of the nicotinic acetylcholine receptor. *Phys. Biol.* **3**, 147–159 (2006).
  29. Wang, H., Cheng, X., Taylor, P., McCammon, J. & Sine, S. Control of cation permeation through the nicotinic receptor channel. *PLoS Comput. Biol.* **4**, e41 (2008).
  30. Corringer, P. J. *et al.* Mutational analysis of the charge selectivity filter of the alpha7 nicotinic acetylcholine receptor. *Neuron* **22**, 831–843 (1999).
  31. Thompson, A. J. & Lummis, S. C. R. A single ring of charged amino acids at one end of the pore can control ion selectivity in the 5-HT<sub>3</sub> receptor. *Br. J. Pharmacol.* **140**, 359–365 (2003).
  32. Cymes, G. D. & Grosman, C. The unanticipated complexity of the selectivity-filter glutamates of nicotinic receptors. *Nature Chem. Biol.* **8**, 975–981 (2012).
  33. Zuber, B. & Unwin, N. Structure and superorganization of acetylcholine receptor-rapsyn complexes. *Proc. Natl Acad. Sci. USA* **110**, 10622–10627 (2013).
  34. Bouzat, C., Bren, N. & Sine, S. M. Structural basis of the different gating kinetics of fetal and adult acetylcholine receptors. *Neuron* **13**, 1395–1402 (1994).
  35. Kelley, S. P., Dunlop, J. I., Kirkness, E. F., Lambert, J. J. & Peters, J. A. A cytoplasmic region determines single-channel conductance in 5-HT<sub>3</sub> receptors. *Nature* **424**, 321–324 (2003).
  36. Peters, J. A. *et al.* Novel structural determinants of single channel conductance and ion selectivity in 5-hydroxytryptamine type 3 and nicotinic acetylcholine receptors. *J. Physiol. (Lond.)* **588**, 587–596 (2010).
  37. McKinnon, N., Bali, M. & Akabas, M. H. 5-HT<sub>3</sub> receptor ion size selectivity is a property of the transmembrane channel, not the cytoplasmic vestibule portals. *J. Gen. Physiol.* **138**, 453–466 (2011).
  38. Carland, J. E. *et al.* Mutagenic analysis of the intracellular portals of the human 5-HT<sub>3A</sub> receptor. *J. Biol. Chem.* **288**, 31592–31601 (2013).
  39. Kozuska, J. L. *et al.* Impact of intracellular domain flexibility upon properties of activated human 5-HT<sub>3</sub> receptors. *Br. J. Pharmacol.* **171**, 1617–1628 (2014).
  40. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
  41. Hansen, S. B. *et al.* Structures of Aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24**, 3635–3646 (2005).
  42. Kesters, D. *et al.* Structural basis of ligand recognition in 5-HT<sub>3</sub> receptors. *EMBO Rep.* **14**, 49–56 (2013).
  43. Huang, S. *et al.* Complex between  $\alpha$ -bungarotoxin and an  $\alpha$ 7 nicotinic receptor ligand-binding domain chimera. *Biochem. J.* **454**, 303–310 (2013).
  44. Beene, D. L. *et al.* Cation- $\pi$  interactions in ligand recognition by serotonergic (5-HT<sub>3A</sub>) and nicotinic acetylcholine receptors: the anomalous binding properties of nicotine. *Biochemistry* **41**, 10262–10269 (2002).
  45. Zhong, W. *et al.* From ab initio quantum mechanics to molecular neurobiology: A cation- $\pi$  binding site in the nicotinic receptor. *Proc. Natl Acad. Sci. USA* **95**, 12088–12093 (1998).
  46. Miles, T. F., Bower, K. S., Lester, H. A. & Dougherty, D. A. A coupled array of noncovalent interactions impacts the function of the 5-HT<sub>3A</sub> serotonin receptor in an agonist-specific way. *ACS Chem. Neurosci.* **3**, 753–760 (2012).
  47. Hu, X.-Q., Zhang, L., Stewart, R. R. & Weight, F. F. Arginine 222 in the pre-transmembrane domain 1 of 5-HT<sub>3A</sub> receptors links agonist binding to channel gating. *J. Biol. Chem.* **278**, 46583–46589 (2003).
  48. Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).
  49. Mukhtasimova, N. & Sine, S. M. Nicotinic receptor transduction zone: invariant arginine couples to multiple electron-rich residues. *Biophys. J.* **104**, 355–367 (2013).
  50. Li, T. *et al.* Cell-penetrating anti-GFAP VHH and corresponding fluorescent fusion protein VHH-GFP spontaneously cross the blood-brain barrier and specifically recognize astrocytes: application to brain imaging. *FASEB J.* **26**, 3969–3979 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Research in the laboratory of H.V. was supported by the Swiss National Science Foundation (grant 31003A-133141; SystemsX/CINA program), internal funds of the Ecole Polytechnique Fédérale de Lausanne, the European Community's Seventh Framework Program FP7/2007-2013 under Grant 211800, the Human Frontier Science Program (postdoc fellowship to H.N.). Additional support was provided by The Agence Nationale de la Recherche (VenomPicoScreen project, grant ANR-11-RPIB-022-04). We thank S. Thurnheer and the expression facility, F. Pojer and the crystallization facility at the EPFL, D. Blot, the HTXLab and the SLS and ESRF beamlines staff. We are particularly appreciative of discussions with H.V. laboratory members and E. Pebay-Peyroula. We thank the company TheraNyx for the discovery and production of the functional nanobody used for the structure resolution.

**Author Contributions** H.V. initiated the project and was responsible for overall project management and strategy. G.H. designed the research for the expression, purification and nanobody aspects. C.D., G.H. and H.N. performed purification. A.D. and G.H. made the nanobodies. C.D., G.H., H.N., L.G., R.W., M.B.T., R.H. and C.M. performed biochemical, biophysical and functional characterization of the receptor and/or of the complexes with nanobodies. A.G. and H.S. performed electron microscopy characterization. X.-D.L., T.T., C.D. and H.N. performed crystallization and data collection. H.N. solved the structure. H.N. and F.P. analysed the structure. H.N. wrote the manuscript with help of H.V. and F.P. All authors discussed results and commented on manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Coordinates have been deposited in the Protein Data Bank under accession number 4PIR. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.N. ([hugues.nury@ibs.fr](mailto:hugues.nury@ibs.fr)) or H.V. ([horst.vogel@epfl.ch](mailto:horst.vogel@epfl.ch)).

# Interacting supernovae from photoionization-confined shells around red supergiant stars

Jonathan Mackey<sup>1</sup>, Shazrene Mohamed<sup>2</sup>, Vasilii V. Gvaramadze<sup>3,4,5</sup>, Rubina Kotak<sup>6</sup>, Norbert Langer<sup>1</sup>, Dominique M.-A. Meyer<sup>1</sup>, Takashi J. Moriya<sup>1</sup> & Hilding R. Neilson<sup>7</sup>

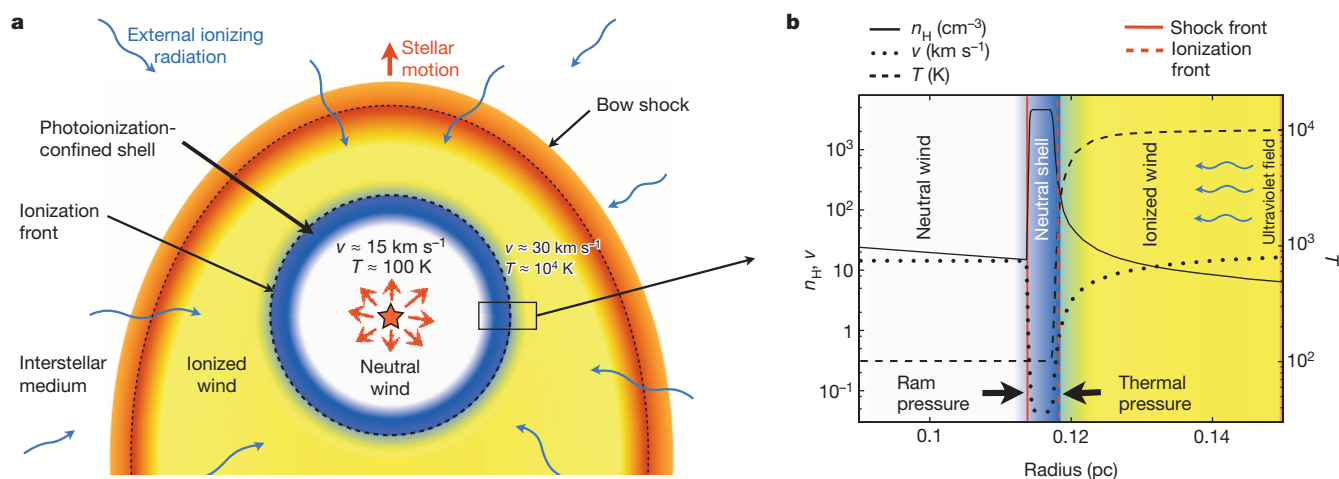
Betelgeuse, a nearby red supergiant, is a fast-moving star with a powerful stellar wind that drives a bow shock into its surroundings<sup>1–4</sup>. This picture has been challenged by the discovery of a dense and almost static shell<sup>5</sup> that is three times closer to the star than the bow shock and has been decelerated by some external force. The two physically distinct structures cannot both be formed by the hydrodynamic interaction of the wind with the interstellar medium. Here we report that a model in which Betelgeuse's wind is photoionized by radiation from external sources can explain the static shell without requiring a new understanding of the bow shock. Pressure from the photoionized wind generates a standing shock in the neutral part of the wind<sup>6</sup> and forms an almost static, photoionization-confined shell. Other red supergiants should have much more massive shells than Betelgeuse, because the photoionization-confined shell traps up to 35 per cent of all mass lost during the red supergiant phase, confining this gas close to the star until it explodes. After the supernova explosion, massive shells dramatically affect the supernova light curve, providing a natural explanation for the many supernovae that have signatures of circumstellar interaction.

Red supergiants are massive stars near the end of their lives, and are direct progenitors of core-collapse supernovae<sup>7,8</sup>. They evolve from O- and B-type stars (hot, luminous sources of ionizing photons), and so these stars are often found together, within or near star clusters<sup>9</sup>. As a result, the cool stellar winds of red supergiants are often photoionized by external radiation fields<sup>10–13</sup>. To calculate the radiation hydrodynamics

of a photoionized red supergiant wind, we simplify the problem by assuming spherical symmetry. We use an approximate two-temperature equation of state for the gas, for which both the neutral gas and the photoionized gas are isothermal with temperatures  $T = T_n$  and  $T_i \gg T_n$ , respectively. The ionized and neutral isothermal sound speeds similarly satisfy  $a_i \gg a_n$ . The photoionized part of the red supergiant wind is accelerated as a result of ionization heating<sup>14</sup>, whereas the neutral part is decelerated<sup>6</sup> if the wind speed through the ionization front,  $v_n$ , satisfies  $v_n \leq 2a_i$ .

The resulting flow is depicted in Fig. 1. The outermost layer is the interface where the wind meets the interstellar medium. For static stars this is a spherical, detached shell, and for stars moving supersonically it is a bow shock. A photoionization-confined shell—a dense, shocked layer separating the neutral inner wind from the ionized outer wind—forms closer to the star. We identify this with the recently discovered shell in Betelgeuse's circumstellar medium<sup>5</sup>.

The properties of the photoionization-confined shell are calculated analytically and verified with simulations in Methods. Its outer boundary,  $R_{IF}$ , is calculated following previous work<sup>10</sup> (Extended Data Fig. 1), and the standing shock radius,  $R_{shell}$ , is obtained by requiring pressure balance across the shell. The shell reaches its final position (determined by the wind density and the incoming photon flux) on the expansion timescale of the wind, and then accumulates mass until it reaches a steady state, where the gas added to the shell at  $R_{shell}$  is balanced by that photo-evaporated from  $R_{IF}$ . The steady-state mass of the shell,  $M_{shell}$ , follows from its density and volume (Extended Data Fig. 2). For realistic

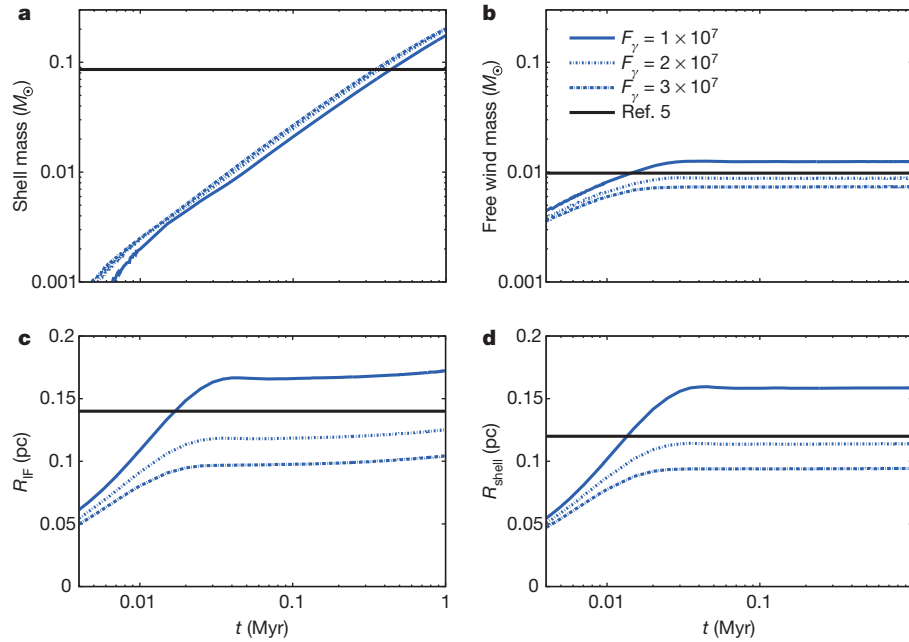


**Figure 1 | Circumstellar structures produced when a runaway red supergiant is exposed to an external ionizing radiation field.** **a**, A neutral stellar wind expands freely from the star and is shocked and decelerated by a photoionization-confined shell. A photoionized wind accelerates away from the shell's outer surface until it reaches the interface between the wind and the

interstellar medium, which is a bow shock for Betelgeuse. **b**, Detailed structure of a photoionization-confined shell from a spherically symmetric radiation hydrodynamics simulation of Betelgeuse's wind. Hydrogen number density,  $n_H$ , velocity,  $v$ , and temperature,  $T$ , are plotted as functions of radius.

<sup>1</sup>Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany. <sup>2</sup>South African Astronomical Observatory, PO Box 9, 7935 Observatory, South Africa. <sup>3</sup>Sternberg Astronomical Institute, Lomonosov Moscow State University, Universitetskij Prospekt 13, Moscow 119992, Russia. <sup>4</sup>Isaac Newton Institute of Chile, Moscow Branch, Universitetskij Prospekt 13, Moscow 119992, Russia. <sup>5</sup>Space Research Institute, Russian Academy of Sciences, Profsoyuznaya 84/32, Moscow 117997, Russia. <sup>6</sup>Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK. <sup>7</sup>Department of Physics and Astronomy, East Tennessee State University, Box 70652, Johnson City, Tennessee 37614, USA.





**Figure 2 | Time evolution of the photoionization-confined shell around Betelgeuse for spherically symmetric simulations with three different ionizing fluxes.** **a**,  $M_{\text{shell}}$ ; **b**, mass in the freely expanding wind,  $M_{\text{wind}}$ ; **c**,  $R_{\text{IF}}$ ; **d**,  $R_{\text{shell}}$ . All calculated using wind parameters  $\dot{M} = 1.2 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$

wind properties and radiation fluxes, the most likely radii and masses are  $R_{\text{IF}} \approx 0.003\text{--}0.3 \text{ pc}$  and  $M_{\text{shell}} \approx (0.03\text{--}10) M_{\odot}$  ( $M_{\odot}$ , solar mass). Photoionization-confined shells are present in addition to bow shocks and detached shells, and should be common because red supergiant winds are often photoionized<sup>10–13</sup>.

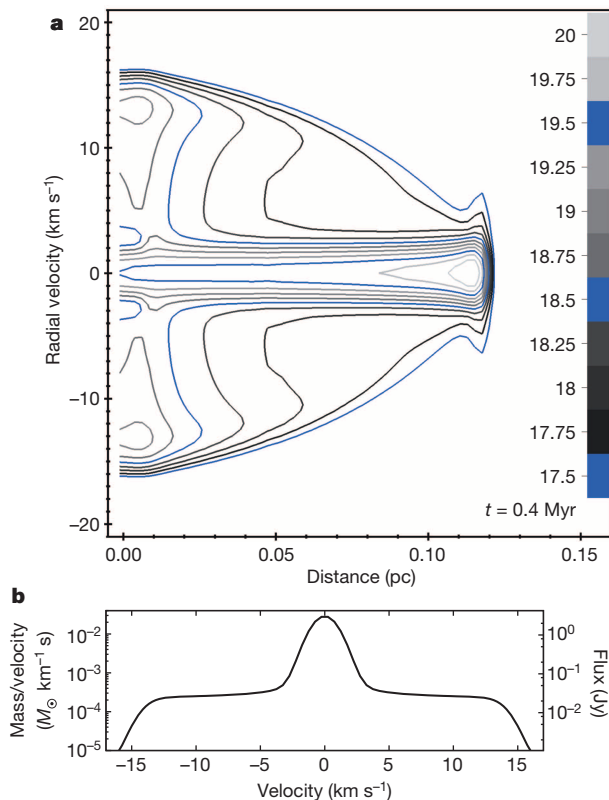
The steady-state shell mass for Betelgeuse is  $M_{\text{shell}} = 1.0 M_{\odot}$ , for the parameter values<sup>5</sup>  $v_n = 14 \text{ km s}^{-1}$ ,  $R_{\text{IF}} \approx 0.15 \text{ pc}$  and  $\dot{M} = 1.2 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$  (stellar mass-loss rate). If  $\dot{M}$  is larger (for example  $3 \times 10^{-6}$

and  $v_n = 14 \text{ km s}^{-1}$ . The blue curves are from simulations with different external ionizing fluxes,  $F_{\gamma} (\text{cm}^{-2} \text{ s}^{-1})$ . The black lines are plotted using data in table 2 of ref. 5 ( $M_{\text{DT,CS}}$ ,  $M_{r < r_1}$ ,  $r_f$  and  $r_1$  and **a–d**, respectively), calibrated to match the 21 cm H I observations.

$M_{\odot} \text{ yr}^{-1}$ ; ref. 3) then  $M_{\text{shell}}$  increases accordingly. In Fig. 2, we compare our model predictions for Betelgeuse to a recent analysis<sup>5</sup> in which 21 cm H I observations were interpreted in the context of a detached shell model. A photoionization-confined shell matches the observations well for an external ionizing flux of  $F_{\gamma} \approx 2 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$ . Such a flux is found near the edge of old H II regions or within interstellar bubbles where diffuse ionizing photons constitute a large fraction of the total flux<sup>15</sup> (Methods; Extended Data Fig. 4 and Supplementary Video 1 show results from this simulation). The observed shell mass ( $0.09 M_{\odot}$ ) constrains its age to be 0.3–0.5 Myr. Betelgeuse's post-main-sequence lifetime is about 1 Myr, and so it will probably explode before the photoionization-confined shell attains its steady-state mass.

Further quantitative comparison at simulation time  $t = 0.4 \text{ Myr}$  is shown in Fig. 3. The H I column density is plotted in a position–velocity diagram as a function of separation from Betelgeuse and line-of-sight velocity of the gas. The blue- and red-shifted components of the freely expanding wind are at radial velocity  $v_r \approx \pm 14 \text{ km s}^{-1}$ , and the shocked shell is centred on  $v_r = 0 \text{ km s}^{-1}$ . Our results again agree quantitatively with the data presented in ref. 5 (fig. 11 therein), with the caveat that our simulations did not self-consistently determine the shell temperature (nor, consequently, the thermal broadening).

Although the neutral shell has been clearly seen observationally, a crucial prediction of our model is that the photoionization-confined shell should be surrounded by an accelerating ionized wind, emitting bremsstrahlung at radio wavelengths (with an emission measure of  $\sim 10\text{--}20 \text{ cm}^{-6} \text{ pc}$ ; emission measure is the integral of the square of the electron



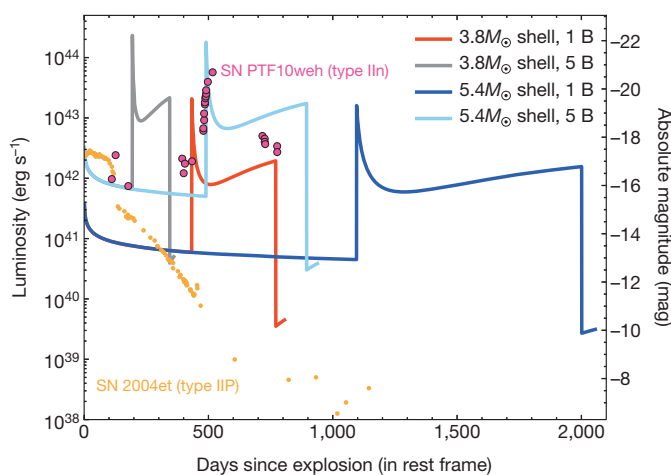
**Figure 3 | Simulated observations of neutral hydrogen in the photoionization-confined shell around Betelgeuse.** Observations are the output of a simulation using the models described in Fig. 2 with  $F_{\gamma} = 2 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$  at  $t = 0.4 \text{ Myr}$ , when  $M_{\text{shell}} = 0.093 M_{\odot}$ . **a**, Position–velocity diagram showing logarithmic contours of H I column density as a function of projected distance from the star and radial velocity, in units of  $\log_{10}[\text{H I atoms } (\text{cm}^{-2} \text{ km}^{-1} \text{ s})]$ . The freely expanding wind is seen red- and blueshifted by  $14 \text{ km s}^{-1}$  (thermally broadened), and the almost static shell is at zero velocity and is limb-brightened at large radius. **b**, Total spectrum of the H I emission, assuming a distance of 200 pc and that the source is unresolved and spherically symmetric (mass of H I per unit velocity also shown).

number density along a line of sight) and Doppler-shifted nebular spectral lines. Such a nebula has been detected for the red supergiant W 26<sup>9,12</sup>, but it will be orders of magnitude fainter around Betelgeuse because the latter's wind has much lower density. The only imaging detections of Betelgeuse's gaseous circumstellar medium so far are the 21 cm H I data and unexplained far-ultraviolet emission from the bow shock<sup>5</sup>.

The agreement of Betelgeuse's neutral shell with our photoionization-confined shell calculations is encouraging, but further work is required to investigate multidimensional effects such as the non-radial flows that are introduced by clumpy and asymmetric winds<sup>16</sup>, the dynamical stability of the shocked shell, and the anisotropic external radiation fields<sup>10</sup> (Methods). Photoionization-confined shells may also be present around lower-mass red giants and stars on the asymptotic giant branch that have winds of comparable velocity, if they are located in a photoionized medium.

The main effect of a photoionization-confined shell is to confine much more gas (up to 80 times more (Methods)) close to a red supergiant than would be expected from a freely expanding wind. Simulations show that ~20–35% of the red supergiant wind is trapped in the photoionization-confined shell (Extended Data Fig. 3). The shell mass is ultimately limited by the total mass shed by the star during the red supergiant phase of evolution, which is typically less than  $20M_{\odot}$  at solar metallicity, and so we expect the most massive shells to have  $M_{\text{shell}} \approx (4 - 7)M_{\odot}$ .

This has important implications for supernova/circumstellar-medium interactions because ejected material from about 10% of all core-collapse supernovae is observed to collide with dense circumstellar matter in the immediate vicinity of the exploding star<sup>7</sup>. It is usually assumed that this dense matter is produced by short periods of extraordinarily high mass-loss rate ( $\lesssim 0.1M_{\odot} \text{ yr}^{-1}$ ) just before the star explodes<sup>16–21</sup>. This gas is difficult to decelerate and confine close to the star hydrodynamically<sup>22</sup>, requiring a prompt explosion after an eruptive mass-loss event. There is, however, no proven evolutionary link between eruptions and explosions (although ideas are being investigated<sup>23,24</sup>). The photoionization-confined shell scenario overcomes this timing problem because the wind is decelerated effectively, allowing a fundamentally different interpretation of the light curves of some interacting supernova. For example, Betelgeuse's mass-loss rate was previously deemed too small to produce an interacting supernova<sup>16</sup>, but this conclusion may need revision following the detection of its photoionization-confined shell<sup>5</sup>.



**Figure 4 | Predicted luminosity evolution of supernovae interacting with massive photoionization-confined shells, compared with observations of two core-collapse supernovae.** The shells are from simulations with  $\dot{M} = 10^{-4}M_{\odot} \text{ yr}^{-1}$ ,  $v_{\infty} = 15 \text{ km s}^{-1}$ , and either  $F_{\gamma} = 10^{14} \text{ cm}^{-2} \text{ s}^{-1}$  ( $3.8M_{\odot}$  shell) or  $F_{\gamma} = 10^{13} \text{ cm}^{-2} \text{ s}^{-1}$  ( $5.4M_{\odot}$  shell at larger radius). The shell grows for 0.2 Myr until the star has lost  $20M_{\odot}$  of mass, and then it explodes. The supernova kinetic energy is either  $10^{51} \text{ erg}$  (equal to 1 Bethe (B)) or  $5 \times 10^{51} \text{ erg}$  (5 B), as indicated. Also plotted are R-band light curves of the interacting supernova PTF10weh (ref. 26) and the more typical supernova 2004et (ref. 25).

Results from a calculation of a radiative supernova blast wave expanding through the circumstellar medium of two of our most extreme models are plotted in Fig. 4. The bolometric light curve rebrillens when the blast wave reaches the photoionization-confined shell, and remains at nearly constant luminosity until the shell is overrun by the shock. The initial interaction with the circumstellar medium is strong enough to leave detectable signatures in supernova observations, and the later shell collision is even more luminous. Two core-collapse supernova light curves are shown for qualitative comparison: SN 2004et (ref. 25) belongs to the most common (plateau) class, whereas SN PTF10weh (ref. 26) is an interacting supernova. The light curve of SN PTF10weh was interpreted as a bright pre-supernova eruption (0–500 d in Fig. 4) followed by a luminous supernova at ~500 d (ref. 26). Figure 4 shows an alternative interpretation that fits the data well: an explosion at 0 d followed by a collision of the supernova shock with a photoionization-confined shell at ~500 d. In this interpretation, the supernova would produce broad spectral lines immediately after explosion. Because pre-supernova eruptions eject lower velocity gas with narrower lines, this could be used to distinguish the two scenarios.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 March; accepted 16 May 2014.

Published online 10 August 2014.

- Noriega-Crespo, A., van Buren, D., Cao, Y. & Dgani, R. A parsec-size bow shock around Betelgeuse. *Astron. J.* **114**, 837–840 (1997).
- Ueta, T. et al. AKARI/FIS mapping of the ISM-wind bow shock around  $\alpha$  Orionis. *Pub. Astron. Soc. Jpn* **60**, S407–S413 (2008).
- Mohamed, S., Mackey, J. & Langer, N. 3D simulations of Betelgeuse's bow shock. *Astron. Astrophys.* **541**, A1 (2012).
- Decin, L. et al. The enigmatic nature of the circumstellar envelope and bow shock surrounding Betelgeuse as revealed by Herschel. I. Evidence of clumps, multiple arcs, and a linear bar-like structure. *Astron. Astrophys.* **548**, A113 (2012).
- Le Bertre, T., Matthews, L. D., Gérard, E. & Libert, Y. Discovery of a detached H I gas shell surrounding  $\alpha$  Orionis. *Mon. Not. R. Astron. Soc.* **422**, 3433–3443 (2012).
- Kahn, F. D. The acceleration of interstellar clouds. *Bull. Astron. Inst. Neth.* **12**, 187–200 (1954).
- Smartt, S. J. Progenitors of core-collapse supernovae. *Annu. Rev. Astron. Astrophys.* **47**, 63–106 (2009).
- Langer, N. Presupernova evolution of massive single and binary stars. *Annu. Rev. Astron. Astrophys.* **50**, 107–164 (2012).
- Dougherty, S. M., Clark, J. S., Negueruela, I., Johnson, T. & Chapman, J. M. Radio emission from the massive stars in the galactic super star cluster Westerlund 1. *Astron. Astrophys.* **511**, A58 (2010).
- Morris, M. & Jura, M. The nature of NML Cygnus. *Astrophys. J.* **267**, 179–183 (1983).
- Yusef-Zadeh, F. & Morris, M. A windswept cometary tail on the Galactic center supergiant IRS 7. *Astrophys. J.* **371**, L59–L62 (1991).
- Wright, N. J. et al. The ionized nebula surrounding the red supergiant W26 in Westerlund 1. *Mon. Not. R. Astron. Soc.* **437**, L1–L5 (2014).
- Gvaramadze, V. V. et al. IRC-10414: a bow-shock-producing red supergiant star. *Mon. Not. R. Astron. Soc.* **437**, 843–856 (2014).
- Meyer, D. M.-A. et al. On the stability of bow shocks generated by red supergiants: the case of IRC-10414. *Mon. Not. R. Astron. Soc.* **439**, L41–L45 (2014).
- Ritzerveld, J. The diffuse nature of Strömgren spheres. *Astron. Astrophys.* **439**, L23–L26 (2005).
- Smith, N., Hinkle, K. H. & Ryde, N. Red supergiants as potential type IIIn supernova progenitors: spatially resolved 4.6  $\mu\text{m}$  CO emission around VY CMa and Betelgeuse. *Astron. J.* **137**, 3558–3573 (2009).
- Smith, N. et al. SN 2006gy: discovery of the most luminous supernova ever recorded, powered by the death of an extremely massive star like  $\eta$  Carinae. *Astrophys. J.* **666**, 1116–1128 (2007).
- Smith, N. et al. Late-time observations of SN 2006gy: still going strong. *Astrophys. J.* **686**, 485–491 (2008).
- Yoon, S.-C. & Cantiello, M. Evolution of massive stars with pulsation-driven superwinds during the red supergiant phase. *Astrophys. J.* **717**, L62–L65 (2010).
- Fox, O. D. et al. Late-time circumstellar interaction in a Spitzer selected sample of type IIIn supernovae. *Astron. J.* **146**, 2 (2013).
- Moriya, T. J. et al. An analytic bolometric light curve model of interaction-powered supernovae and its application to type IIIn supernovae. *Mon. Not. R. Astron. Soc.* **435**, 1520–1535 (2013).
- van Marle, A. J., Langer, N., Achterberg, A. & García-Segura, G. Forming a constant density medium close to long gamma-ray bursts. *Astron. Astrophys.* **460**, 105–116 (2006).
- Quataert, E. & Shiode, J. Wave-driven mass loss in the last year of stellar evolution: setting the stage for the most luminous core-collapse supernovae. *Mon. Not. R. Astron. Soc.* **423**, L92–L96 (2012).



24. Moriya, T. J. Mass loss of massive stars near the Eddington luminosity by core neutrino emission shortly before their explosion. *Astron. Astrophys.* **564**, A83 (2014).
25. Kotak, R. *et al.* Dust and the type II-plateau supernova 2004et. *Astrophys. J.* **704**, 306–323 (2009).
26. Ofek, E. O. *et al.* Precursors prior to type IIin supernova explosions are common: precursor rates, properties, and correlations. *Astrophys. J.* **789**, 104 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** J.M. and S.M. are grateful to P. Kervella, T. Le Bertre and G. Perrin, the organisers of the Betelgeuse Workshop in Paris (November 2012), where the ideas for this work were first developed. J.M. acknowledges funding from a fellowship from the Alexander von Humboldt Foundation and from the Deutsche Forschungsgemeinschaft priority program 1573, 'Physics of the Interstellar Medium'. S.M. acknowledges the receipt of research funding from the National Research Foundation (NRF) of South Africa. T.J.M. is supported by the Japan Society for the Promotion of Science Postdoctoral Fellowships for Research Abroad (26·51). H.R.N. acknowledges funding from a NSF grant (AST-0807664). R.K. acknowledges support

from STFC (ST/L000709/1). The authors acknowledge the John von Neumann Institute for Computing for a grant of computing time on the JUROPA supercomputer at Jülich Supercomputing Centre.

**Author Contributions** J.M. and S.M. had the original idea that Betelgeuse's static shell could be confined by external radiation. J.M. derived the analytic equations for the shell, and ran and analysed the spherically symmetric computations. V.V.G., D.M.-A.M., N.L. and J.M. discussed the results in the context of recently discovered photoionized winds, which motivated many of the specific choices of parameters used. J.M., S.M., V.V.G., D.M.-A.M., H.R.N. and N.L. interpreted Betelgeuse's shell in the context of our results. N.L. proposed that the shells could be relevant for interacting supernovae, and developed this idea with J.M., R.K. and T.J.M. Figures were prepared by J.M., S.M., T.J.M. and R.K. All authors contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. ([jmackey@astro.uni-bonn.de](mailto:jmackey@astro.uni-bonn.de)).

# Magneto-optical trapping of a diatomic molecule

J. F. Barry<sup>1†</sup>, D. J. McCarron<sup>1</sup>, E. B. Norrgard<sup>1</sup>, M. H. Steinecker<sup>1</sup> & D. DeMille<sup>1</sup>

Laser cooling and trapping are central to modern atomic physics. The most used technique in cold-atom physics is the magneto-optical trap (MOT), which combines laser cooling with a restoring force from radiation pressure. For a variety of atomic species, MOTs can capture and cool large numbers of particles to ultracold temperatures (less than  $\sim 1$  millikelvin); this has enabled advances in areas that range from optical clocks to the study of ultracold collisions, while also serving as the ubiquitous starting point for further cooling into the regime of quantum degeneracy. Magneto-optical trapping of molecules could provide a similarly powerful starting point for the study and manipulation of ultracold molecular gases. The additional degrees of freedom associated with the vibration and rotation of molecules, particularly their permanent electric dipole moments, allow a broad array of applications not possible with ultracold atoms<sup>1</sup>. Spurred by these ideas, a variety of methods has been developed to create ultracold molecules. Temperatures below 1 microkelvin have been demonstrated for diatomic molecules assembled from pre-cooled alkali atoms<sup>2,3</sup>, but for the wider range of species amenable to direct cooling and trapping, only recently have temperatures below 100 millikelvin been achieved<sup>4,5</sup>. The complex internal structure of molecules complicates magneto-optical trapping. However, ideas and methods necessary for creating a molecular MOT have been developed<sup>6–11</sup> recently. Here we demonstrate three-dimensional magneto-optical trapping of a diatomic molecule, strontium monofluoride (SrF), at a temperature of approximately 2.5 millikelvin, the lowest yet achieved by direct cooling of a molecule. This method is a straightforward extension of atomic techniques and is expected to be viable for a significant number of diatomic species<sup>6,7</sup>. With further development, we anticipate that this technique may be employed in any number of existing and proposed molecular experiments, in applications ranging from precision measurement<sup>12</sup> to quantum simulation<sup>13</sup> and quantum information<sup>14</sup> to ultracold chemistry<sup>15</sup>.

In laser cooling, the species of interest is illuminated by counter-propagating pairs of laser beams. The laser frequency is detuned just below resonance with an electronic transition. Owing to the Doppler effect, a particle is more likely to absorb a photon from the beam that opposes the particle velocity, slowing the particle. If the subsequent spontaneous emission returns the particle to the same initial state (or another state also excited by the lasers), then this process of absorption and spontaneous emission can be repeated many times (a ‘cycling transition’). This provides a damping force. However, this force does not act to spatially confine the particles. In a MOT, cooling and confinement are produced simultaneously. Three orthogonal pairs of laser beams are spatially overlapped with a quadrupole magnetic field. For a pair of ground and excited state Zeeman sublevels, a deviation from the trap centre generally induces a Zeeman shift that moves the transition closer to or further from resonance with the lasers. For a small deviation along a given laser axis, the transition shifted closest to resonance can be driven by a particular laser polarization; this polarization is chosen for the laser counter-propagating to the direction of the deviation, while the co-propagating laser has the orthogonal polarization. Hence, on average there is a confining force restoring particles towards the centre of the trap.

The SrF MOT described in this work uses techniques very similar to those used for standard atomic MOTs. Anti-Helmholtz coils create a

static quadrupole magnetic field (with an axial field gradient,  $dB_z/dz$ , equal to twice the radial field gradient,  $dB_\rho/d\rho$ ), and pairs of circularly polarized laser beams pass through the centre of this field along three orthogonal axes. The trap is loaded with pulses of SrF from a cryogenic buffer gas beam source<sup>16</sup> that have been slowed using radiation pressure<sup>10</sup>. However, the level structure of SrF dictates that our MOT is similar to a rarely used and poorly understood configuration of atomic MOT (known as type II), which differs in certain characteristics from the most common atomic MOTs (type I).

The type-I MOT employs an  $F \rightarrow F' = F + 1$  cycling transition, where  $F$  is the total angular momentum quantum number and the prime indicates the excited state. For a given polarization, particles in all ground-state Zeeman sublevels are optically coupled to the excited state (all states are ‘bright’). The type-II MOT operates on an  $F \rightarrow F' = F$  or  $F \rightarrow F' = F - 1$  transition, where certain ‘dark’ ground state sublevels are not optically coupled to the excited state. The presence of dark states reduces the spontaneous photon scattering rate; this rate can go to zero in the absence of a mechanism to ‘remix’ dark states with the bright states. Moreover, scattering alone does not ensure a confining force: the scattering rate from the laser counter-propagating to a particle’s deviation from the trap centre must exceed the scattering rate from the laser co-propagating. In type-II systems, the level structure makes it possible for particles to be pumped into a state dark to the counter-propagating laser but bright to the co-propagating laser, so a confining force is not guaranteed<sup>17,18</sup>. Consequently the damping rate and restoring force may be significantly smaller for type-II MOTs than for those of type I. There appears to be no widely accepted understanding of the mechanisms responsible for generating a restoring force in type-II MOTs. Nevertheless, type-II MOTs have been demonstrated in several atomic systems<sup>18–21</sup>; for these, relatively weak confinement and slightly elevated temperature are typically observed. The rotational structure of diatomic molecules generically requires their cycling transitions to correspond to a type-II MOT system; the SrF MOT described here is hence also of type II.

We use a previously demonstrated scheme for creating a cycling transition in SrF (refs 8–10) on the  $X^2\Sigma^+ \rightarrow A^2\Pi_{1/2}$  electronic transition (see Methods). Calculated vibrational branching fractions  $b_{v'v}$ , for decay of the excited state with vibrational quantum number  $v'$  to the ground state with vibrational quantum number  $v$ , suggest that only three vibrationally excited levels in the X state ( $v = 1, 2, 3$ ) are significantly populated after  $\sim 10^6$  photon scattering events, corresponding to  $\sim 1$  s of optical cycling for typical scattering rates (see below). Hence, three vibrational repumping wavelengths are expected to be sufficient to trap SrF for the  $\sim 1$  s timescale typical of atomic MOTs. We denote the laser addressing the  $X(v=i) \rightarrow A(v'=j)$  transition by  $\mathcal{L}_{ij}$ , so that the three repump lasers are labelled  $\mathcal{L}_{10}$ ,  $\mathcal{L}_{21}$  and  $\mathcal{L}_{32}$ ; the primary and secondary trapping lasers are denoted  $\mathcal{L}_{00}$  and  $\mathcal{L}_{00}^\dagger$  respectively. (The need for a second trapping laser is explained below.) Radio-frequency sidebands on the  $\mathcal{L}_{00}$ ,  $\mathcal{L}_{10}$ ,  $\mathcal{L}_{21}$  and  $\mathcal{L}_{32}$  lasers address spin-rotation/hyperfine (SR/HF) substructure in the  $X^2\Sigma$  state (Fig. 1a). Rotational branching is eliminated by driving an  $N = 1(J = 3/2, 1/2) \rightarrow J' = 1/2$  transition<sup>7</sup>, where  $N$  is the total angular momentum excluding electronic and nuclear spin and  $J$  is the total angular momentum excluding nuclear spin. Driving these transitions optically pumps population into dark ground-state

<sup>1</sup>Department of Physics, Yale University, PO Box 208120, New Haven, Connecticut 06520, USA. <sup>†</sup>Present address: Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.



Zeeman sublevels not excited by the laser<sup>22</sup>. These dark states must be remixed with the bright states for cycling to continue. In this work, remixing occurs both due to Larmor precession in the quadrupole magnetic field and due to optical pumping as molecules move through the complicated optical polarization gradients arising from the orthogonal pairs of circularly polarized laser beams<sup>23</sup>.

The optimal polarization of the trapping light depends both on the sign of the difference in magnetic moment between the ground and excited states of the cycling transition and on the orientation of the quadrupole magnetic field. Within the sublevels of the SrF  $X^2\Sigma(N=1)$  state, two of the four SR/HF manifolds have positive magnetic  $g$ -factors, one has  $g=0$ , and the remaining manifold has  $g<0$  (Fig. 1a); the  $A^2\Pi_{1/2}$  state has  $g\approx 0$ . The presence of both negative and positive  $g$ -factors means that laser frequencies addressing the different SR/HF manifolds must have different polarizations for optimal trapping. This is achieved by combining the  $\mathcal{L}_{00}$  light with single-frequency light of the opposite polarization from the  $\mathcal{L}_{00}^\dagger$  secondary trapping laser (see Methods).

The pulse of molecules from the beam source begins with laser ablation of an SrF<sub>2</sub> target at  $t=0$  ms. The slowing is applied from  $t=0$  ms to  $t=40$  ms (see Methods). Molecules in the trapping region are detected via laser-induced fluorescence (LIF) from the  $X\rightarrow A$  cycling transition at  $\lambda_{00}=663.3$  nm and imaged onto a CCD (Fig. 1b; see Methods).

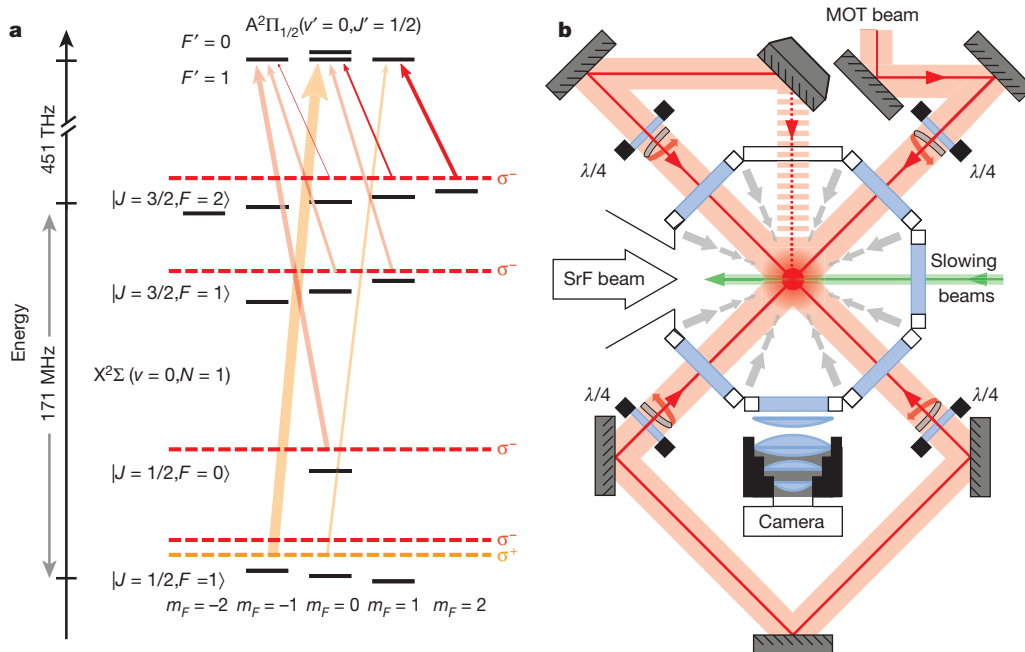
Realization of magneto-optical trapping results in increased LIF from a small area in the trapping region near the  $B$ -field zero. This localized LIF persists for an increased duration compared to the spatially broad LIF from the untrapped, slowed molecular beam and suggests that molecules are confined in this region. To observe the MOT, the  $\mathcal{L}_{00}$ ,  $\mathcal{L}_{00}^\dagger$ ,  $\mathcal{L}_{10}$  and  $\mathcal{L}_{21}$  lasers must be present with the proper detunings (and, for the trapping lasers, polarizations), the  $B$ -field gradient must be present ( $dB_z/dz\neq 0$ ) and the laser slowing must be applied. The  $\mathcal{L}_{32}$  laser is not necessary to observe the MOT but results in increased LIF. Maximum LIF is observed with the laser detunings  $\Delta_{00}=\Delta_{00}^\dagger=-1.2\Gamma\approx 2\pi\times 8$  MHz (where  $\Gamma=2\pi\times 7$  MHz is the natural linewidth) and  $dB_z/dz=15\text{ G cm}^{-1}$  (see Methods); these parameters are similar to those for

standard atomic MOTs. Unless stated otherwise, measurements are made with these default parameters.

The proper polarization for the trapping light depends on the sign of  $dB_z/dz$ . Reversing either the sign of  $dB_z/dz$  or the circular polarization of the MOT trapping light should create an anti-restoring force and prevent MOT formation. Reversing both the sign of  $dB_z/dz$  and the polarization in tandem, however, should return the system to a restoring configuration, and the MOT should be realized again. We observe the expected behaviour for these four states of the system as shown in Fig. 2, confirming magneto-optical trapping of SrF molecules. From these images we also determine the MOT cloud position and size by fitting the LIF intensity profile to a two-dimensional Gaussian; we find typical r.m.s. widths of  $\rho_{\text{rms}}=4.1(1)$  mm (radial) and  $z_{\text{rms}}=2.6(1)$  mm (axial).

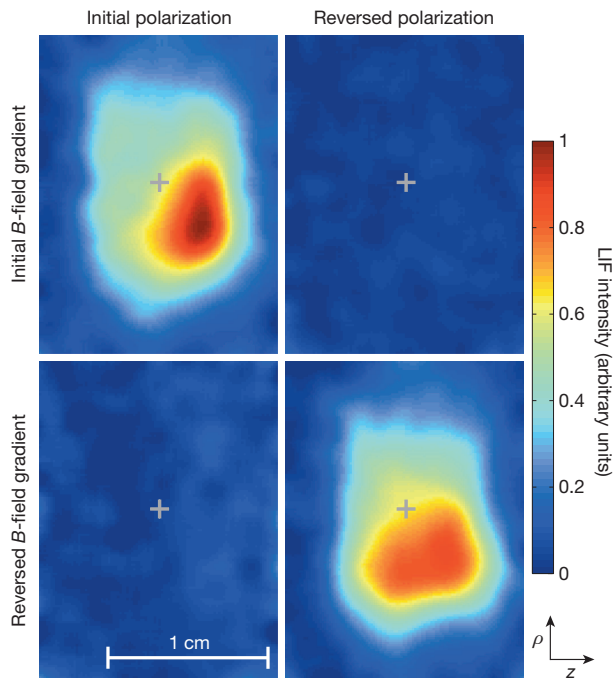
To probe the confining and damping forces in the MOT, the molecular cloud's response to a rapid displacement of the trap centre is measured. During loading, a magnetic bias field offsets the MOT centre radially, along the axis of the molecular beam. The bias field is then switched off, releasing the trapped molecules into the unbiased potential. Using a short camera exposure ( $\Delta t_{\text{exp}}=5$  ms), the molecular cloud's position is measured as a function of time (Fig. 3a; see Methods). The cloud exhibits damped harmonic motion as it moves towards the equilibrium position, with oscillation frequency  $\omega_\rho=2\pi\times 17.2(6)$  Hz and damping coefficient  $\alpha/m_{\text{SrF}}=140(10)\text{ s}^{-1}$ , where  $m_{\text{SrF}}$  is the mass. With the measured radial width ( $\rho_{\text{rms}}$ ), the equipartition theorem is used to find the radial MOT temperature:  $T_\rho=\omega_\rho^2\rho_{\text{rms}}^2/k_B m_{\text{SrF}}=2.5(2)$  mK, where  $k_B$  is the Boltzmann constant. If we assume the relation  $\omega_z^2=2\omega_\rho^2$  holds, as for a standard atomic MOT in a quadrupole field<sup>24</sup>, we find the axial oscillation frequency  $\omega_z=2\pi\times 24.3(9)$  Hz. The measured MOT axial width  $z_{\text{rms}}$  then corresponds to an axial temperature  $T_z=2.0(1)$  mK.

To verify the MOT temperature, ballistic expansion measurements are performed. Trapped molecules are released at  $t_{\text{rel}}=90$  ms (by extinguishing the  $\mathcal{L}_{21}$  laser). After a time of flight  $\Delta t_{\text{TOF}}$ , the  $\mathcal{L}_{21}$  light is restored, and the resulting LIF is imaged onto the CCD. A short imaging time ( $\Delta t_{\text{exp}}=5$  ms) is used to accurately determine the expanded cloud's size



**Figure 1 | Experimental set-up.** **a**, Optical addressing scheme. Relevant energy levels are shown for a positive  $B$ -field;  $m_F$  is the angular momentum projection. The  $\mathcal{L}_{00}$  laser (red dashed lines) primarily addresses the  $|J=3/2, F=2\rangle$ ,  $|J=3/2, F=1\rangle$  and  $|J=1/2, F=0\rangle$  states, while the  $\mathcal{L}_{00}^\dagger$  laser (orange dashed line) addresses the  $|J=1/2, F=1\rangle$  state. Arrows show  $\sigma^-$  transitions (red) driven by the  $\mathcal{L}_{00}$  laser light and  $\sigma^+$  transitions (orange) driven by the  $\mathcal{L}_{00}^\dagger$  laser light. Transitions from  $|J=3/2, F=1\rangle$ ,  $|J=1/2, F=0\rangle$  and

$|J=1/2, F=1\rangle$  are marked with transparent arrows for clarity; each line width is proportional to the transition strength. The lasers are drawn at the ground state energy with which they would be resonant. **b**, Experiment schematic showing the MOT (red) and slowing (green) laser beam paths. The line widths indicate beam diameters, and the grey arrows illustrate the default magnetic field gradient. The waveplates and mirrors used to create the vertical MOT beam (dashed line) are not shown.

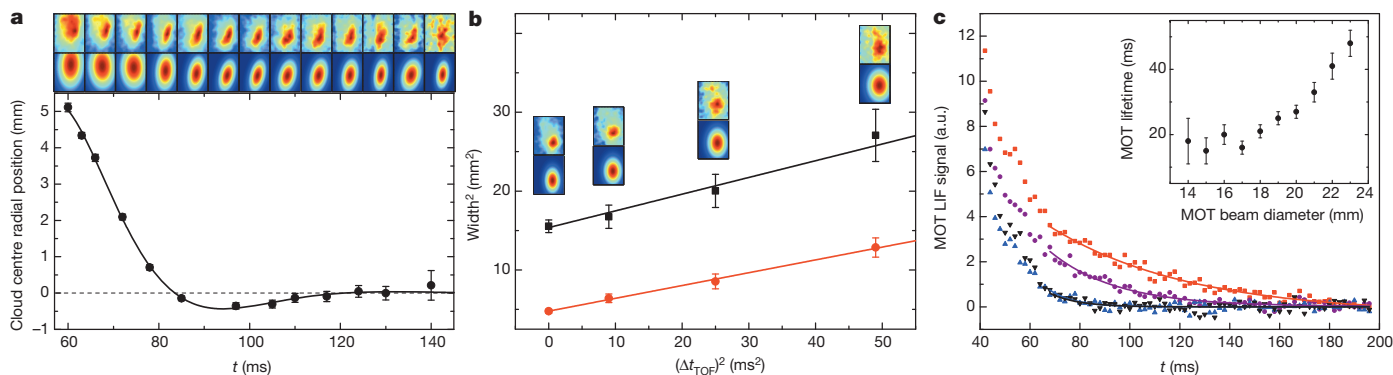


**Figure 2 | Magneto-optical trapping of SrF.** Shown are images (averaged over 600 pulses) of LIF in the trapping region for different polarizations and signs of  $dB_z/dz$ . The grey crosses mark the position of the magnetic field zero. Gravity is in the  $z$ -direction. For clarity, data in the images are smoothed with a Gaussian of width  $\sigma = 0.7$  mm.

(see Methods). For an initial Gaussian spatial distribution and a Boltzmann distribution of velocities (with no correlation between position and velocity), the widths  $z_{\text{rms}}$  and  $\rho_{\text{rms}}$  of the expanding cloud are given by:

$$\rho_{\text{rms}}^2 = \frac{k_B T_\rho}{m_{\text{SrF}}} \left( \frac{1}{\omega_\rho^2} + (\Delta t_{\text{TOF}})^2 \right) \text{ and } z_{\text{rms}}^2 = \frac{k_B T_z}{m_{\text{SrF}}} \left( \frac{1}{\omega_z^2} + (\Delta t_{\text{TOF}})^2 \right)$$

The data and associated fits are shown in Fig. 3b. The slopes of the fits give the temperatures, which are then used with the intercepts to determine  $\omega_\rho$  and  $\omega_z$ . This treatment plots the measured cloud width



**Figure 3 | Measurement of MOT properties.** **a**, MOT cloud response to rapid displacement of the trap centre. Top, LIF images averaged over 1,600 pulses (images are placed at the appropriate position on the  $x$  axis); middle, 2D Gaussian fits to LIF images; bottom (main panel), the extracted radial position as a function of time. The fit is to the motion of a damped harmonic oscillator. Zero is set at the position of the MOT with no displacement. **b**, Free expansion of the MOT following release. For a given position on the  $x$  axis (free expansion time), insets show LIF images averaged over 2,000 pulses (top) and 2D Gaussian fits (bottom), and data points show measured MOT radial (black

(which is an average over the camera exposure duration) at the start time of imaging/illumination. Monte Carlo simulations for the measured trap frequencies suggest that the cloud continues to expand during the short illumination interval, and therefore the extracted width is an upper bound for the actual width at the imaging time,  $t_{\text{im}} = t_{\text{rel}} + \Delta t_{\text{TOF}}$ . Hence this treatment of the data yields upper limits on the MOT temperature. The fits give  $T_\rho \leq 2.7(3)$  mK and  $T_z \leq 2.1(1)$  mK, with trap frequencies  $\omega_\rho = 2\pi \times 19(1)$  Hz and  $\omega_z = 2\pi \times 29(1)$  Hz, corresponding to spring constants  $\kappa_\rho = 2.5(3) \times 10^{-21}$  N m $^{-1}$  and  $\kappa_z = 5.9(4) \times 10^{-21}$  N m $^{-1}$ . These values are in good agreement with the values from the MOT oscillation measurement. The temperatures are roughly an order of magnitude greater than the SrF Doppler temperature,  $T_D = \hbar I / (2k_B) = 160$   $\mu$ K. Temperatures well above the Doppler temperature are also reported in work with atomic type-II MOTs $^{21}$ . These spring constants are two to three orders of magnitude smaller than for typical type-I atomic MOTs $^{25}$  and approximately one order of magnitude smaller than reported values for atomic type-II MOTs $^{21}$ , though measurements of the spring constants in type-II atomic MOTs are so few that ‘typical’ behaviour is difficult to characterize. A third measurement of the MOT temperature  $T_{\text{MOT}}$  is performed using the release-and-recapture method, yielding a temperature in good agreement with the other methods (see Methods).

Measurement of the spontaneous photon scattering rate for trapped molecules,  $R_{\text{sc}}$ , allows the number of trapped molecules  $N_{\text{MOT}}$  to be determined via fluorescence detection. We find  $R_{\text{sc}} \approx 4.3_{-2.2}^{+4.1} \times 10^6$  s $^{-1}$  (see Methods). Based on the efficiency of the LIF detection system, measured to be  $\sim 0.8\%$ , the MOT population is estimated at  $N_{\text{MOT}} \approx 300$  SrF molecules, corresponding to a peak trap density of  $n_{\text{MOT}} \approx 600$  cm $^{-3}$ .

The MOT lifetime,  $\tau_{\text{MOT}}$ , is obtained by measuring LIF in the trapping region as a function of time and fitting a single exponential decay curve to the data after  $t = 67$  ms, as shown in Fig. 3c. This start time avoids significant contributions to the LIF signal from the slowed but ultimately untrapped part of the molecular beam. We find  $\tau_{\text{MOT}} = 56(4)$  ms, significantly shorter than is typically seen in atomic MOTs. When the  $\mathcal{L}_{32}$  repump laser is not present,  $\tau_{\text{MOT}} = 27(2)$  ms. We have verified that neither collisions with ballistic helium from the buffer-gas beam nor collisions with background gases are the primary loss mechanism from the trap. Optical pumping into the dark  $X^2\Sigma(v=4)$  state would result in  $\tau_{\text{MOT}} \approx 1$  s for the measured value of  $R_{\text{sc}}$  and off-resonant excitation populating dark rotational levels is found to be insignificant (see Methods).

The strikingly low restoring force measured suggests another explanation for the low value of  $\tau_{\text{MOT}}$ : the trap depth is not large compared to  $k_B T_{\text{MOT}}$ , as in typical atomic MOTs, so a significant fraction of molecules

squares) and axial (red circles) widths. In both **a** and **b**, images are rescaled to the maximum value at each time. **c**, Main panel; LIF in the trapping region versus time  $t$  for MOT with (red squares) and without (purple circles) the  $\mathcal{L}_{32}$  repump laser, for untrapped damping (blue triangles), and for damping/anti-restoring (black triangles) configurations (see Methods). Overlaid are single exponential fits. Inset, MOT lifetime versus MOT laser beam diameter. Error bars,  $\pm 1\sigma$  confidence interval from a  $\chi^2$  analysis of the fits. LIF images are smoothed as in Fig. 2; all fits and analyses are performed using the unsmoothed data.



can escape the trap simply by being in the high-energy tail of the Boltzmann distribution. The MOT trap depth  $U_{\text{MOT}}$  can be estimated using  $U_{\text{MOT}} = \frac{1}{2} \kappa_{\rho} (d_{\text{L}}/2)^2$ , assuming that  $\kappa_{\rho}$  is constant to the edges of the MOT beam ( $d_{\text{L}}$  is the beam diameter). This gives  $U_{\text{MOT}}/k_{\text{B}} = 10(1) \text{ mK} \approx 4T_{\text{MOT}}$ , in contrast to atomic MOTs where  $U_{\text{MOT}}/k_{\text{B}} \approx 1 \text{ K} \approx 1,000T_{\text{MOT}}$ . We presume that rapid molecule–light interactions maintain a constant temperature in the MOT, leading to continuous loss rather than evaporative cooling as in a conservative trap (where the trap leaves the total energy of a trapped sample unchanged). A simple model for the rate of particle escape under these conditions lends credence to this explanation for the short MOT lifetime (see Methods). Additional support comes from the observation that  $\tau_{\text{MOT}}$  depends strongly on the MOT beam diameter (Fig. 3c, inset). Reducing the beam diameter  $d_{\text{L}}$  from 23 mm to 21 mm (a <1% decrease in power) reduces  $\tau_{\text{MOT}}$  by ~30%. We are unaware of any other trap loss mechanism that might exhibit this behaviour.

For our cycling transition, the maximum restoring force  $F_{\text{max}} = \kappa_{\text{z}} d_{\text{L}}/2$  corresponds to a confining photon scattering rate  $R_{\text{con}} = F_{\text{max}}/(h\kappa) = 5(2) \times 10^4 \text{ s}^{-1}$  (where  $k = 2\pi/\lambda_{00}$  is the wavenumber) confining photons from a single MOT beam, only ~1% of  $R_{\text{sc}}$ . The small value of  $R_{\text{con}}/R_{\text{sc}}$  may be understood qualitatively by noting that in a simple one-dimensional model, the angular momentum level structure of our system ( $J = 3/2, 1/2 \rightarrow J' = 1/2$ ) ensures that each photon scattered in the ‘correct’ (confining) direction on average must be followed by a photon scattered in the ‘incorrect’ (anti-confining) direction in order to resume optical cycling<sup>18</sup>. In three dimensions, with complicated polarization gradients and other means of remixing, this relation no longer holds exactly. Nonetheless, the mechanism behind the slight bias of scattering events towards the trap centre that leads to the weak, yet non-zero confining force is not well understood. This same type of level structure is also the defining characteristic of atomic type-II MOTs, which exhibit qualitatively similar characteristics to our SrF MOT (such as extended spatial extent and elevated temperature) although with reported stronger confinement<sup>19–21</sup>. Hence the weak trapping and only moderately low temperature observed in our SrF MOT are believed to be due to the angular momentum level structure rather than any other issues related specifically to using a diatomic molecule rather than an atom. Despite these limitations, our method succeeds in trapping and cooling molecules to the lowest temperature reported for any direct-cooling method to date.

Future work is expected to allow substantial increases in the density and the number of molecules trapped. For example, the trappable flux may be increased by implementing a more efficient slowing method<sup>26</sup> or by transversely confining the molecular beam as it is slowed<sup>27</sup>. A variety of methods may enable increased trap depth by increasing the fraction of scattered photons contributing to the confining force ( $R_{\text{con}}/R_{\text{sc}}$ ), which could in turn increase trap lifetime, capture velocity, density, and number of molecules trapped. This could be accomplished, for example, by rapid synchronous reversing of the MOT magnetic field gradient and the laser circular polarizations, as recently demonstrated in two-dimensional magneto-optic compression of a molecular beam<sup>11</sup>, or alternatively by using a microwave electric field to pump molecules in anti-trapped Zeeman sub-levels to trapped levels by driving transitions through other rotational states<sup>7,8</sup>.

Although magneto-optical trapping of molecules is in its infancy, our results demonstrate that this technique could be applied in a straightforward way to a significant number of diatomic species<sup>6,7</sup>. The MOT has proved indispensable for cooling and trapping many atomic species; with further development, we expect that it may prove similarly useful for producing ultracold gases of diatomic molecules. Such an advance is expected to enable a wide range of new experiments including tests of the standard model of particle physics<sup>28,29</sup>, sensitive searches for variations of fundamental constants<sup>30</sup>, and studies of novel chemical dynamics in the ultracold temperature regime<sup>7</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 April; accepted 23 June 2014.

- Carr, L., DeMille, D., Krems, R. & Ye, J. Cold and ultracold molecules: science, technology and applications. *New J. Phys.* **11**, 055049 (2009).
- Ni, K.-K. *et al.* A high phase-space-density gas of polar molecules. *Science* **322**, 231–235 (2008).
- Danzl, J. G. *et al.* An ultracold high-density sample of rovibronic ground-state molecules in an optical lattice. *Nature Phys.* **6**, 265–270 (2010).
- Zeppenfeld, M. *et al.* Sisyphus cooling of electrically trapped polyatomic molecules. *Nature* **491**, 570–573 (2012).
- Stuhl, B. K. *et al.* Evaporative cooling of the dipolar hydroxyl radical. *Nature* **492**, 396–400 (2012).
- Di Rosa, M. D. Laser-cooling molecules. *Eur. Phys. J. D* **31**, 395–402 (2004).
- Stuhl, B. K., Sawyer, B. C., Wang, D. & Ye, J. Magneto-optical trap for polar molecules. *Phys. Rev. Lett.* **101**, 243002 (2008).
- Shuman, E. S., Barry, J. F., Glenn, D. R. & DeMille, D. Radiative force from optical cycling on a diatomic molecule. *Phys. Rev. Lett.* **103**, 223001 (2009).
- Shuman, E. S., Barry, J. F. & DeMille, D. Laser cooling of a diatomic molecule. *Nature* **467**, 820–823 (2010).
- Barry, J. F., Shuman, E. S., Norrgard, E. B. & DeMille, D. Laser radiation pressure slowing of a molecular beam. *Phys. Rev. Lett.* **108**, 103002 (2012).
- Hummon, M. T. *et al.* 2D magneto-optical trapping of diatomic molecules. *Phys. Rev. Lett.* **110**, 143001 (2013).
- The ACME Collaboration. Order of magnitude smaller limit on the electric dipole moment of the electron. *Science* **343**, 269–272 (2014).
- Micheli, A., Brennen, G. K. & Zoller, P. A toolbox for lattice-spin models with polar molecules. *Nature Phys.* **2**, 341–347 (2006).
- DeMille, D. Quantum computation with trapped polar molecules. *Phys. Rev. Lett.* **88**, 067901 (2002).
- Krems, R. V. Cold controlled chemistry. *Phys. Chem. Chem. Phys.* **10**, 4079–4092 (2008).
- Hutzler, N. R., Lu, H.-I. & Doyle, J. M. The buffer gas beam: an intense, cold, and slow source for atoms and molecules. *Chem. Rev.* **112**, 4803–4827 (2012).
- Prentiss, M. G., Bigelow, N. P., Shahriar, M. S. & Hemmer, P. R. Forces on three-level atoms including coherent population trapping. *Opt. Lett.* **16**, 1695–1697 (1991).
- Nasyrov, K. *et al.* Magneto-optical trap operating on a magnetically induced level-mixing effect. *Phys. Rev. A* **64**, 023412 (2001).
- Prentiss, M., Cable, A., Bjorkholm, J. E., Chu, S. & Raab, E. L. Atomic-density-dependent losses in an optical trap. *Opt. Lett.* **13**, 452–454 (1988).
- Flemming, J. *et al.* Magneto-optical trap for sodium atoms operating on the D<sub>1</sub> line. *Opt. Commun.* **135**, 269–272 (1997).
- Tiwari, V. B., Singh, S., Rawat, H. S. & Mehendale, S. C. Cooling and trapping of <sup>85</sup>Rb atoms in the ground hyperfine F=2 state. *Phys. Rev. A* **78**, 063421 (2008).
- Berkeland, D. J. & Boshier, M. G. Destabilization of dark states and optical spectroscopy in Zeeman-degenerate atomic systems. *Phys. Rev. A* **65**, 033413 (2002).
- Rio Fernandes, D. *et al.* Sub-Doppler laser cooling of fermionic <sup>40</sup>K atoms in three-dimensional gray optical molasses. *Europhys. Lett.* **100**, 63001 (2012).
- Metcalf, H. J. & van der Straten, P. *Laser Cooling and Trapping* (Springer, 1999).
- Wallace, C. D. *et al.* Measurements of temperature and spring constant in a magneto-optical trap. *J. Opt. Soc. Am. B* **11**, 703–711 (1994).
- Chieda, M. A. & Eyer, E. E. Prospects for rapid deceleration of small molecules by optical bichromatic forces. *Phys. Rev. A* **84**, 063401 (2011).
- DeMille, D., Barry, J. F., Edwards, E. R., Norrgard, E. B. & Steinecker, M. H. On the transverse confinement of radiatively slowed molecular beams. *Mol. Phys.* **111**, 1805–1813 (2013).
- Hunter, L. R., Peck, S. K., Greenspon, A. S., Alam, S. S. & DeMille, D. Prospects for laser cooling TIF. *Phys. Rev. A* **85**, 012511 (2012).
- Tarbutt, M. R., Sauer, B. E., Hudson, J. J. & Hinds, E. A. Design for a fountain of YbF molecules to measure the electron’s electric dipole moment. *New J. Phys.* **15**, 053034 (2013).
- Chin, C., Flambaum, V. V. & Kozlov, M. G. Ultracold molecules: new probes on the variation of fundamental constants. *New J. Phys.* **11**, 055048 (2009).

**Acknowledgements** We thank E.R. Edwards for contributions towards the construction of the experiment. We acknowledge funding from AFOSR (MURI), ARO, and ARO (MURI). E.B.N. acknowledges funding from the NSF GRFP.

**Author Contributions** All authors contributed to the experiment, the analysis of the results and the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J.M. ([daniel.mccarron@yale.edu](mailto:daniel.mccarron@yale.edu)).

# Abrupt glacial climate shifts controlled by ice sheet changes

Xu Zhang<sup>1</sup>, Gerrit Lohmann<sup>1,2</sup>, Gregor Knorr<sup>1,3</sup> & Conor Purcell<sup>1,3</sup>

During glacial periods of the Late Pleistocene, an abundance of proxy data demonstrates the existence of large and repeated millennial-scale warming episodes, known as Dansgaard–Oeschger (DO) events<sup>1</sup>. This ubiquitous feature of rapid glacial climate change can be extended back as far as 800,000 years before present (BP) in the ice core record<sup>2</sup>, and has drawn broad attention within the science and policy-making communities alike<sup>3</sup>. Many studies have been dedicated to investigating the underlying causes of these changes, but no coherent mechanism has yet been identified<sup>3–15</sup>. Here we show, by using a comprehensive fully coupled model<sup>16</sup>, that gradual changes in the height of the Northern Hemisphere ice sheets (NHISs) can alter the coupled atmosphere–ocean system and cause rapid glacial climate shifts closely resembling DO events. The simulated global climate responses—including abrupt warming in the North Atlantic, a northward shift of the tropical rainbelts, and Southern Hemisphere cooling related to the bipolar seesaw—are generally consistent with empirical evidence<sup>1,3,17</sup>. As a result of the coexistence of two glacial ocean circulation states at intermediate heights of the ice sheets, minor changes in the height of the NHISs and the amount of atmospheric CO<sub>2</sub> can trigger the rapid climate transitions via a local positive atmosphere–ocean–sea-ice feedback in the North Atlantic. Our results, although based on a single model, thus provide a coherent concept for understanding the recorded millennial-scale variability and abrupt climate changes in the coupled atmosphere–ocean system, as well as their linkages to the volume of the intermediate ice sheets during glacials.

A common explanation for DO cycles involves changes in the Atlantic meridional overturning circulation<sup>5,6</sup> (AMOC), perhaps triggered by freshwater forcing<sup>3–6</sup>. However, the representation of freshwater origin, timing and magnitude—used to force abrupt ocean circulation changes—varies between models, and palaeoceanographic evidence for these ocean circulation changes and their relationship to freshwater perturbation remains elusive (Methods). Nevertheless, one fundamental characteristic of DO events, possibly offering a clue to their origin, is that almost all events occurred during glacial periods when global ice volume was varying at intermediate levels<sup>1,2,18</sup>, suggesting a potential relationship between the intermediate ice sheets and the existence of millennial-scale climate variability<sup>7</sup>. To test this hypothesis, we use a fully coupled Earth System Model (COSMOS)<sup>16</sup> to assess how changes in the height of the NHISs affect global climate during glacial periods, and find that the greater height of the NHISs forces a relatively stronger AMOC. As discussed later, a nonlinear behaviour of the AMOC exists under the intermediate height of the NHISs, equivalent to a decrease in sea level of ~60 m relative to today (Methods).

To investigate the characteristics of this abrupt climate shift, we conduct a transient simulation (ISTran45) by gradually increasing the height of the NHISs from that of the intermediate ice sheets, at a rate of decrease of 1.8 cm yr<sup>-1</sup> equivalent sea level (ESL) (Fig. 1a–f). In response to a linear increase in ice sheets, surface air temperature (SAT) in the northern North Atlantic is characterized by a gradual warming rather than by a two-step strengthening process demonstrated by the AMOC strength (Fig. 1a–c). Initially, over the first 70 years, SAT in the northern North

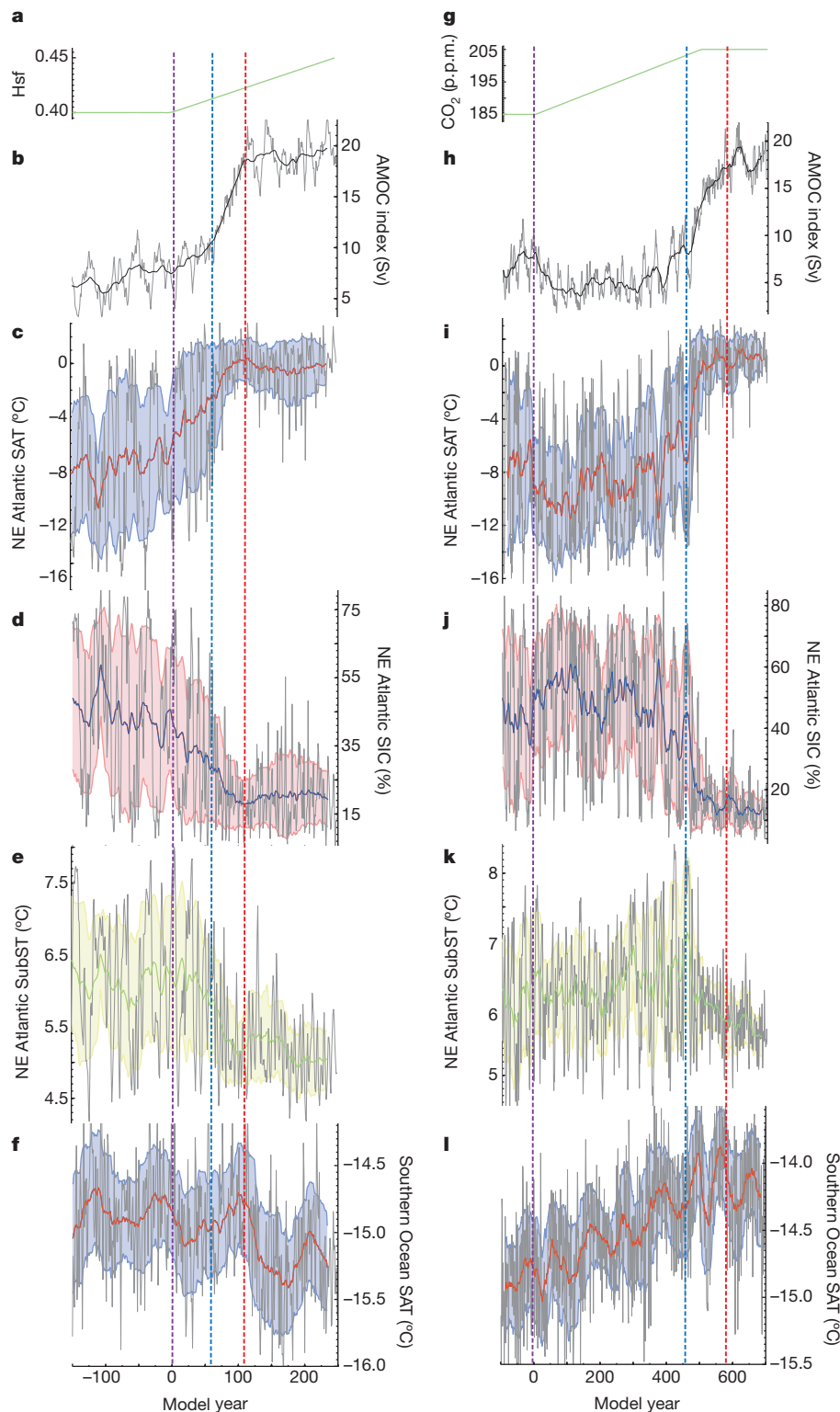
Atlantic gradually warms up by 4 °C, coincident with an increase in AMOC strength of 3 sverdrups (Sv, where 1 Sv is equal to a volume flow rate of 10<sup>6</sup> m<sup>3</sup> s<sup>-1</sup>). Once the SAT warming has passed a thermal threshold (about -2.5 °C in our model), the weak AMOC increases abruptly by ~10 Sv within half a century, attaining a strong overturning mode, along with a further warming of ~3 °C in the northern North Atlantic (Figs 1b, c and 2a, b).

The SAT in the northern North Atlantic is characterized by strong variability during the weak AMOC mode (Fig. 1c, model years -150 to 0), also accompanied by variability in sea-ice concentration (SIC) and subsurface temperature (SubST) (Figs 1c–e and 2c). In fact, the SAT, SIC and SubST variabilities of the warm climate state (strong AMOC mode) lie within the range of high variability of the cold climate state (weak AMOC mode) (Fig. 1c–e and Extended Data Figs 3, 4). We attribute this high variability associated with the weak AMOC mode to the interplay between sea-ice change and subsurface warming in the northern North Atlantic. Indeed, the cold climate (weak AMOC mode) supports sea-ice formation and the build-up of a strong halocline in the northern North Atlantic. However, under the intermediate height of the ice sheets, the accumulated warm subsurface water mass can gradually destabilize the water column, triggering the release of subsurface heat and resuming a warm sea surface temporally<sup>19,20</sup> (Fig. 1c–e). Once the gradual surface warming in the northern North Atlantic passes the thermal threshold, the reduced SIC and the weakened stratification undermine the interaction between the sea-ice change and subsurface warming, lowering the internal climate variability (for example SAT, SIC and SubST) (Figs 1c–e and 2c, d). Thus, the surface warming acts to increase air–sea heat flux by reducing SIC, promoting a fast attainment of vigorous deep convection. This further warms the northern North Atlantic and completes the abrupt transition to a warm climate state (Fig. 1b–d). In parallel, the high variability in the northern North Atlantic undergoes a shift, along with the mode transition to the Nordic Sea, causing a fluctuation of SIC and resulting in large temperature variations there (Fig. 2c, d and Extended Data Fig. 4).

The abrupt transition from weak to strong AMOC modes is controlled by a nonlinear response of the glacial ocean to gradual variations in the NHIS height by two mechanisms: first, changes in the export of sea ice to the North Atlantic, and second, adjustments of the subpolar and subtropical gyre systems, both of which occur in response to changes in the wind field forced by variations in the NHIS height, especially the Laurentide Ice Sheet (LIS; Extended Data Figs 2 and 5). Our model simulations show that variations in the height of the LIS lead to changes in the gyre circulation and sea-ice coverage by shifting the Northern westerlies. Associated with a gradual increase of the NHISs, the maximum westerly wind stress shifts northwards to ~57° N. Consequently, the zonal wind stress over the South Labrador Sea decreases, effectively weakening the export of sea ice to the northeastern North Atlantic. This reduces the SIC and results in surface warming over the region, permitting open ocean convection and the formation of North Atlantic Deep Water. In parallel, a strengthened wind-stress curl accelerates the North Atlantic gyre systems, encouraging both the northwards advection of heat via the

<sup>1</sup>Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bussestrasse 24, D-27570 Bremerhaven, Germany. <sup>2</sup>MARUM-Center for Marine Environmental Sciences, University Bremen, Leobener Strasse, D-28359 Bremen, Germany. <sup>3</sup>School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3AT, UK.



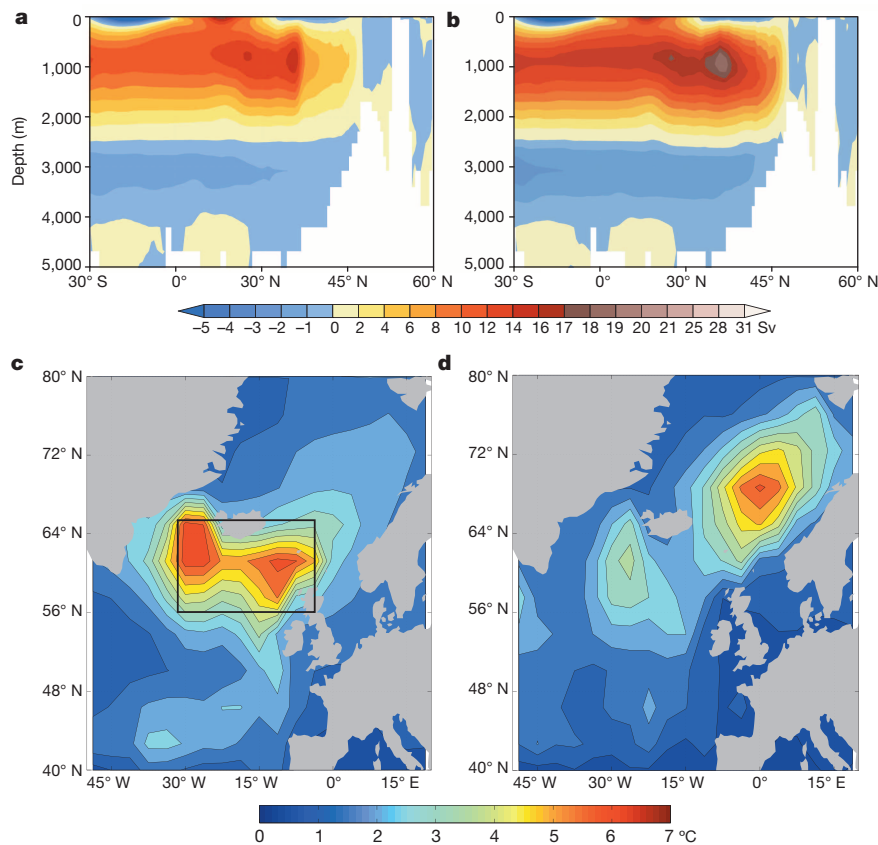


**Figure 1 | Transient simulations with gradually increasing NHIS height (ISTran45) and CO<sub>2</sub> level (TrGHG04).** **a, g,** Transient forcing; **b, h,** AMOC index (AMOC maximum in the North Atlantic north of 45° N; **c, i,** SAT index; **d, j,** SIC index; **e, k,** SubST index (at 100–1,000 m water depth) of the northern North Atlantic (average in 56–65° N, 5–30° W); **f, l,** SAT index in the Southern Ocean (60–76° S zonal mean). **a–f,** transient experiment ISTran45; **g–l,** transient experiment TrGHG04. Bold lines show the 30-year running mean of the original data (grey lines). Shadings indicate one standard deviation of the indices in 30-year windows. The vertical purple, blue and red dotted lines represent the starting points for the transient simulations, abrupt AMOC transitions and cooling in the Southern Hemisphere, respectively. Negative model years indicate the control simulation of NHIS\_0.4 (Extended Data Table 1). The mode transition in ISTran45 occurs in about model year 60 (blue dashed line in **a**). This transition shows that a decrease of ~1.1 m in ESL (corresponding to a 1.2% increase in NHIS height) can trigger the rapid glacial climate change when the NHISs are close to its height threshold. Hsf, height scaling factor (Methods).

North Atlantic Current and vertical mixing in the subpolar North Atlantic, the latter acting to increase the ventilation of subsurface warm waters, resulting in large-scale heat loss and further convection in the open ocean.

To gain a deeper understanding on the effect of variability in ice-sheet height on millennial-scale glacial climate and the governing mechanisms of such change, the stability of the AMOC with respect to intermediate NHIS heights was further tested in COSMOS (Fig. 3 and Methods). After the hysteresis, branch 1 (points a and b in Fig. 3a, b) demonstrates the abrupt increase in AMOC, branches 2 (points b–e in Fig. 3a, b) and 3 (points e and f in Fig. 3a, b) represent the warm glacial climate state and

rapid Northern Hemisphere cooling, respectively, and branch 4 (points f–h in Fig. 3a, b) represents the cold glacial climate state. Indeed, the associated change in AMOC after this hysteresis curve is remarkably similar to the general shape of the DO cycle<sup>1</sup> (Fig. 3b). Most notably, the AMOC hysteresis curve demonstrates a glacial ocean characterized by a classical bistable regime, corresponding to changes in sea-level of up to ~18 m as indicated by proxy data<sup>21</sup> (Fig. 3a). Within the bistable range, two distinct glacial ocean modes with strong and weak AMOC (not an ‘off’ mode) coexist under identical boundary conditions and are characterized by spatial patterns, for example abrupt warming in the North Atlantic, a



**Figure 2 | Ocean circulation and internal SAT variability under cold and warm climate states in experiment IStran45.** Cold and warm climate defined as model years  $-200$  to  $0$  and  $110$ – $250$  of IStran45 (Fig. 1a and Methods). **a, c**, AMOC (**a**) and standard deviation of SAT (**c**) in the cold climate. **b, d**, AMOC (**b**) and standard deviation of SAT (**d**) for the warm climate.

northward shift of the tropical rain belts, and Southern Hemisphere cooling related to the bipolar seesaw, consistent with the existing proxy data for interstadial and stadial conditions, respectively<sup>17</sup> (Figs 1a–f and 3c and Extended Data Fig. 6). In particular, the weak modes associated with the AMOC hysteresis remain characterized by active vertical convection in the North Atlantic (Fig. 2a, b and Extended Data Fig. 2), consistent with proxy data suggesting that major changes in ocean circulation are identified only for the DO–Heinrich events, not for other stadials as depicted here<sup>22</sup>.

Given the same ice-sheet configuration, the strong AMOC mode, in comparison with the weak circulation, is characterized by a reduced SIC in the northern North Atlantic, allowing heat loss to the atmosphere and enhancing the AMOC by the formation of North Atlantic Deep Water (Fig. 4a). In parallel, the warmer SAT associated with this reduced SIC acts to enhance the local atmospheric low-pressure system (Fig. 4b), which in turn boosts the Gulf Stream water mass and heat transport by the altered wind system over the North Atlantic ocean (Fig. 4c). This process represents an atmosphere–ocean–sea-ice positive feedback mechanism stimulating further melting of seaice and thus heat flux from the ocean, which acts to maintain convection in the North Atlantic.

Our results demonstrate that, in combination with the positive atmosphere–ocean–sea-ice feedback, only a slight variation in the height of the intermediate ice sheets (that is, a change of less than 2 m in sea level) is capable of instigating the abrupt AMOC mode transitions by an adjustment of the heat distribution and SIC in the North Atlantic (Figs 1c–e and 3). Although the hydrological compensation due to the global ice volume changes is not considered here, salinification or dilution in the North Atlantic shows a similar effect to increasing or decreasing the NHISs on triggering the mode shifts of the AMOC (Extended

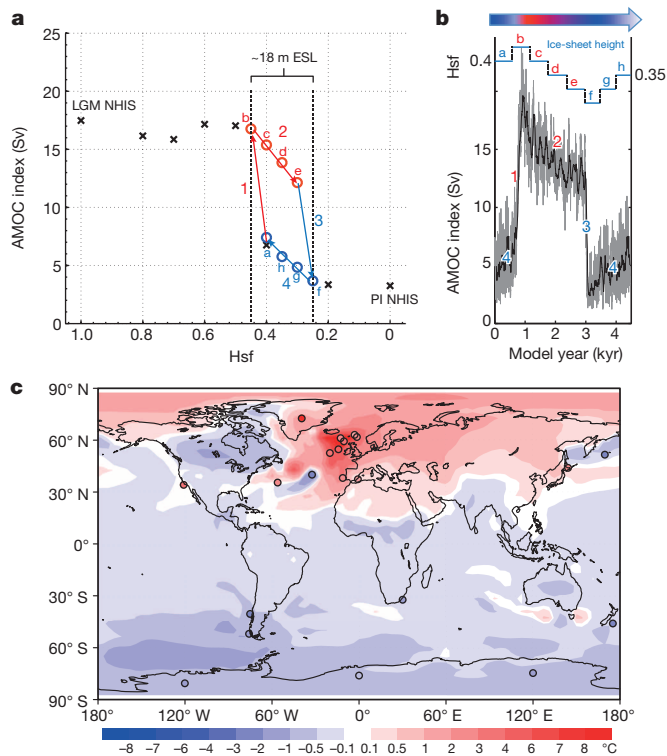
AMOC plots indicate zonally integrated meridional transport in the Atlantic Basin from  $30^{\circ}$  S. Note the increased strength and northward shift of the deep-water convection sites in **b** compared with that in **a**. The black rectangle in **c** indicates the area used to calculate the North Atlantic SAT, SIC and SubST indices in Fig. 1.

Data Fig. 7). However, we attributed the existence of the bistable regime to the positive local atmosphere–ocean feedback mechanism rather than the hydrological balance in the Atlantic basin as suggested by numerous other studies (see, for example, refs 3, 4, 6).

It is known that some large DO warmings, after Heinrich events, are accompanied by large millennial-scale increases in atmospheric carbon dioxide ( $\sim 20$  p.p.m.)<sup>23</sup>, with an impact on a global scale<sup>24</sup>. We have therefore performed a transient experiment (TrGHG04) in which a linear increase in atmospheric  $\text{CO}_2$  concentration ranging from 185 to 205 p.p.m. is imposed over 500 years (at a constant ice-sheet size of 40% maximum glacial level; Methods). In this simulation, a similar yet even more abrupt Northern Hemisphere warming is simulated (Fig. 1g–i). This is due to a release of additional subsurface heat (resulting from a weakened AMOC that exists in response to the increasing  $\text{CO}_2$  level)<sup>25</sup> to the atmosphere, thereafter triggering an abrupt warming of  $\sim 7^{\circ}\text{C}$  in the northern North Atlantic within 20 years. The timing and magnitude of this warming are consistent with proxy data revealing Heinrich–DO events to be simultaneous with increased global  $\text{CO}_2$  (refs 1, 23, 24), indicating that increases in atmospheric  $\text{CO}_2$  alone can also trigger<sup>13</sup> and even amplify the abrupt Northern Hemisphere warming under the intermediate heights of the NHISs.

The model results discussed here demonstrate for the first time that the glacial ocean is characterized by a hysteresis behaviour with respect to changes in the NHISs, corroborating a previous hypothesis that the millennial-scale variability during glacial periods is dependent on the existence of continental ice sheets<sup>7</sup>. The bistable ocean regime accounting for the glacial abrupt climate switches is related to the intermediate volumes of the NHISs, consistent with the fundamental characteristic of the millennial-scale variability in the records<sup>1,2,17,18</sup>. Our results therefore



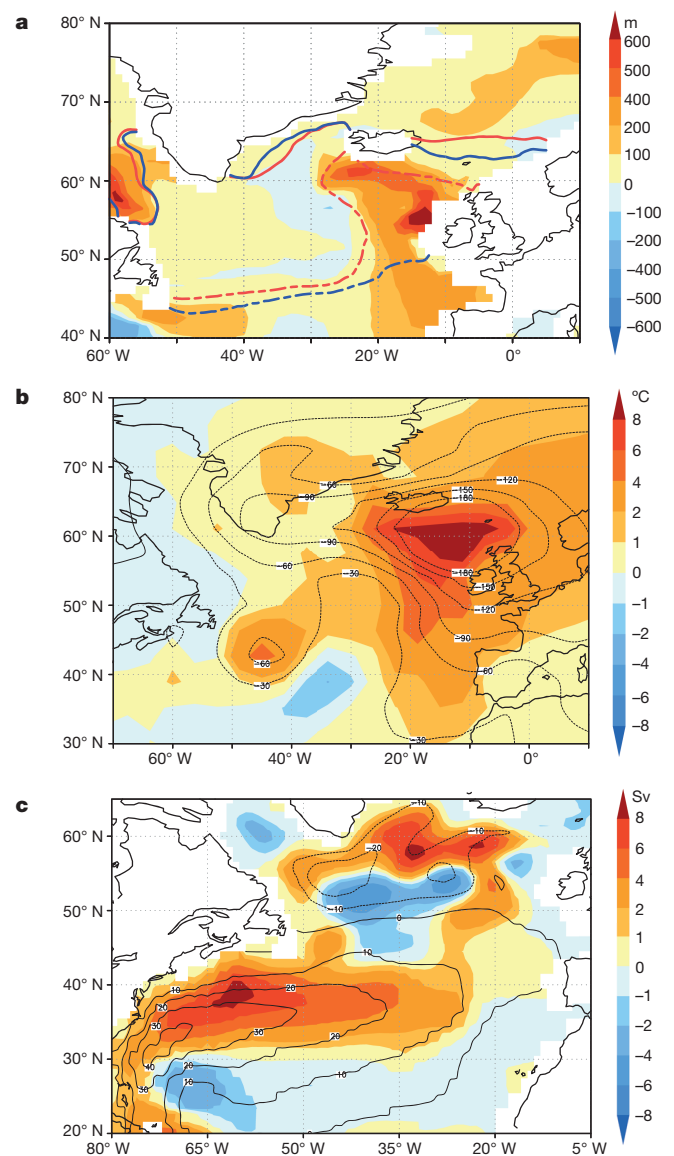


**Figure 3 | AMOC hysteresis with respect to changes in NHIS height and its relationship to recorded abrupt climate variability.** **a**, Response of glacial ocean to variations in the NHIS heights. **b**, AMOC index for the experiments associated with the AMOC hysteresis (Methods). **c**, Simulated annual mean SAT anomaly between the strong and weak AMOC modes with reconstructed temperature changes (circles; Methods). The strong (or weak) mode in **c** and in Fig. 4 is an ensemble mean model results for points b–e (or f–h, **a**) of the hysteresis curve in **a**. Numbers in **b** correspond to the hysteresis branches defined in **a**. The AMOC strength in the pre-industrial (PI) control run is  $\sim 16$  Sv in our model<sup>16</sup>. The low AMOC indices corresponding to the weak modes do not represent an ‘off’ AMOC state (as indicated by the freshwater-hosing-induced Heinrich mode<sup>6,16</sup>), but rather a weakened AMOC (Fig. 2 and Methods). Crosses in **a** represent the experiments performed to analyse the response of the glacial ocean to NHIS change; circles indicate the simulations related to the hysteresis behaviour of glacial ocean (Methods). The NHIS height difference between the Last Glacial Maximum (LGM) and the present-day level is equivalent to a change in sea level of  $\sim 92$  m, indicating an ESL difference between points b and f of about 18 m.

provide a novel and broad framework for understanding the occurrence of rapid climate changes during glacials, indicating that only minor changes in NHISs, atmospheric carbon dioxide and/or hydrological balance in the North Atlantic may induce abrupt ocean circulation change at the intermediate heights of the ice sheets.

The timing of millennial-scale fluctuations in sea level during stadials and interstadials is heavily debated<sup>21,26</sup>. The AMOC hysteresis presented here suggests that a transition from a weak (or strong) to strong (or weak) AMOC mode is compatible with a decrease (or increase) in sea level at the end of stadials (or interstadials). Furthermore, our study indicates a key tempo-spatial change in internal climate variability during rapid glacial climate changes, providing a new dynamic framework to explain the recorded sea surface warming and its increased fluctuation during ice-rafting or at the end of stadials in the northern North Atlantic<sup>27,28</sup>. This characteristic can potentially reconcile the large uncertainty of sea surface reconstructions in the northern North Atlantic and Nordic Sea during glacials<sup>27–29</sup>. Nevertheless, new high-resolution records using improved age models are required for testing these hypotheses further.

Having additionally tested the response of the climate system to changing orbital configurations, we found no significant response in AMOC in our model (Methods). However, the ice sheets in our study are prescribed



**Figure 4 | Annual mean anomaly maps between strong and weak AMOC modes.** **a**, Mixed-layer depth anomaly (shaded) and absolute values for 15% (dashed lines) and 90% (solid lines) sea ice concentration in the strong (red line) and weak (blue line) modes. **b**, **c**, Anomalies of SAT (shaded) and sea-level pressure (contour) (**b**) and anomaly of barotropic horizontal stream function (shaded) and its absolute value in the strong AMOC mode (contour) (**c**).

according to the Last Glacial Maximum configuration and are thus decoupled from changes in external forcing and the internal atmosphere–ocean system. Further studies with similar climate models, preferably including dynamic interactive ice sheet components, should be able to test whether rapid changes in glacial climate can be triggered by variability in ice sheets related to internal feedbacks within atmosphere–ocean–cryosphere systems<sup>15,30</sup> and to weak external forcing<sup>9</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 January; accepted 10 June 2014.

Published online 13 August 2014.

1. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).

2. Barker, S. *et al.* 800,000 years of abrupt climate variability. *Science* **334**, 347–351 (2011).
3. Masson-Delmotte, V. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 383–464 (Cambridge Univ. Press, 2013).
4. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
5. Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. The role of the thermohaline circulation in abrupt climate change. *Nature* **415**, 863–869 (2002).
6. Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature* **419**, 207–214 (2002).
7. Wunsch, C. Abrupt climate change: an alternative view. *Quat. Res.* **65**, 191–203 (2006).
8. Knorr, G. & Lohmann, G. Rapid transitions in the Atlantic thermohaline circulation triggered by global warming and meltwater during the last deglaciation. *Geochim. Geophys. Res.* **12**, Q12006 (2007).
9. Dima, M. & Lohmann, G. Conceptual model for millennial climate variability: a possible combined solar–thermohaline circulation origin for the ~1,500-year cycle. *Clim. Dyn.* **32**, 301–311 (2008).
10. Kwasniok, F. & Lohmann, G. Deriving dynamical models from paleoclimatic records: application to glacial millennial-scale climate variability. *Phys. Rev. E* **80**, 066104 (2009).
11. Arzel, O., Colin de Verdière, A. & England, M. H. The role of oceanic heat transport and wind stress forcing in abrupt millennial-scale climate transitions. *J. Clim.* **23**, 2233–2256 (2010).
12. Li, C., Battisti, D. S. & Bitz, C. M. Can North Atlantic sea ice anomalies account for Dansgaard–Oeschger climate signals? *J. Clim.* **23**, 5457–5475 (2010).
13. Banderas, R., Alvarez-Solas, J. & Montoya, M. Role of CO<sub>2</sub> and Southern Ocean winds in glacial abrupt climate change. *Clim. Past* **8**, 1011–1021 (2012).
14. Dokken, T. M., Nisancioglu, K. H., Li, C., Battisti, D. S. & Kissel, C. Dansgaard–Oeschger cycles: interactions between ocean and sea ice intrinsic to the Nordic seas. *Paleoceanography* **28**, 491–502 (2013).
15. Petersen, S. V., Schrag, D. P. & Clark, P. U. A new mechanism for Dansgaard–Oeschger cycles. *Paleoceanography* **28**, 24–30 (2013).
16. Zhang, X., Lohmann, G., Knorr, G. & Xu, X. Different ocean states and transient characteristics in Last Glacial Maximum simulations and implications for deglaciation. *Clim. Past* **9**, 2319–2333 (2013).
17. Voelker, A. H. L. Global distribution of centennial-scale records for Marine Isotope Stage (MIS) 3: a database. *Quat. Sci. Rev.* **21**, 1185–1212 (2002).
18. Schulz, M., Berger, W. H., Sarinthein, M. & Grootes, P. M. Amplitude variations of 1470-year climate oscillations during the last 100,000 years linked to fluctuations of continental ice mass. *Geophys. Res. Lett.* **26**, 3385–3388 (1999).
19. Rasmussen, T. L. & Thomsen, E. The role of the North Atlantic Drift in the millennial timescale glacial climate fluctuations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **210**, 101–116 (2004).
20. Kim, J.-H. *et al.* Pronounced subsurface cooling of North Atlantic waters off Northwest Africa during Dansgaard–Oeschger interstadials. *Earth Planet. Sci. Lett.* **339–340**, 95–102 (2012).
21. Siddall, M., Rohling, E. J., Thompson, W. G. & Waelbroeck, C. Marine Isotope Stage 3 sea level fluctuations: data synthesis and new outlook. *Rev. Geophys.* **46**, RG000226 (2008).
22. Peck, V. L., Hall, I. R., Zahn, R. & Scourse, J. D. Progressive reduction in NE Atlantic intermediate water ventilation prior to Heinrich events: response to NW European ice sheet instabilities? *Geochim. Geophys. Res.* **8**, Q01N10 (2007).
23. Ahn, J. & Brook, E. J. Atmospheric CO<sub>2</sub> and climate on millennial time scales during the last glacial period. *Science* **83**, 83–85 (2008).
24. Barker, S. & Knorr, G. Antarctic climate signature in the Greenland ice core record. *Proc. Natl Acad. Sci. USA* **104**, 17278–17282 (2007).
25. Gregory, J. M. *et al.* A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO<sub>2</sub> concentration. *Geophys. Res. Lett.* **32**, L12703 (2005).
26. Arz, H. W., Lamy, F., Ganopolski, A., Nowaczyk, N. & Pätzold, J. Dominant Northern Hemisphere climate control over millennial-scale glacial sea-level variability. *Quat. Sci. Rev.* **26**, 312–321 (2007).
27. Peck, V. L., Hall, I. R., Zahn, R. & Elderfield, H. Millennial-scale surface and subsurface paleothermometry from the northeast Atlantic, 55–8 ka BP. *Paleoceanography* **23**, 1–11 (2008).
28. Jonkers, L. *et al.* A reconstruction of sea surface warming in the northern North Atlantic during MIS 3 ice-rafting events. *Quat. Sci. Rev.* **29**, 1791–1800 (2010).
29. Waelbroeck, C. *et al.* Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geosci.* **2**, 127–132 (2009).
30. Macayeal, D. R. Binge/purge oscillations of the Laurentide Ice Sheet as a cause of the North Atlantic's Heinrich events. *Paleoceanography* **8**, 775–784 (1993).

**Acknowledgements** We thank the colleagues in the Paleoclimate Dynamics group at the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI) in Bremerhaven for useful discussions, and I. Hall for comments on a previous draft of the manuscript. G.K. acknowledges helpful discussion with S. Barker. We thank the AWI colleagues who provide technical support on graphics and keep the AWI supercomputer running. This study is promoted by Helmholtz funding through the Polar Regions and Coasts in the Changing Earth System (PACES) programme of the AWI. Funding by 'Helmholtz Climate Initiative REKLIM' (Regional Climate Change), a joint research project of the Helmholtz Association of German research centres (HGF), is gratefully acknowledged (G.K.).

**Author Contributions** X.Z., G.L. and G.K. developed the research and designed the experiments. X.Z. conducted the model simulations and analysed the data. X.Z., G.L. and G.K. interpreted the results. X.Z. led the write-up of the manuscript with significant contribution by C.P. and the other authors. All authors discussed the results and contributed to the preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.Z. ([xu.zhang@awi.de](mailto:xu.zhang@awi.de)).



# Continuing megathrust earthquake potential in Chile after the 2014 Iquique earthquake

Gavin P. Hayes<sup>1</sup>, Matthew W. Herman<sup>2</sup>, William D. Barnhart<sup>1</sup>, Kevin P. Furlong<sup>2</sup>, Sebastián Riquelme<sup>3</sup>, Harley M. Benz<sup>1</sup>, Eric Bergman<sup>4</sup>, Sergio Barrientos<sup>3</sup>, Paul S. Earle<sup>1</sup> & Sergey Samsonov<sup>5</sup>

**The seismic gap theory<sup>1</sup> identifies regions of elevated hazard based on a lack of recent seismicity in comparison with other portions of a fault. It has successfully explained past earthquakes (see, for example, ref. 2) and is useful for qualitatively describing where large earthquakes might occur. A large earthquake had been expected in the subduction zone adjacent to northern Chile<sup>3–6</sup>, which had not ruptured in a megathrust earthquake since a  $M \sim 8.8$  event in 1877. On 1 April 2014 a  $M 8.2$  earthquake occurred within this seismic gap. Here we present an assessment of the seismotectonics of the March–April 2014 Iquique sequence, including analyses of earthquake relocations, moment tensors, finite fault models, moment deficit calculations and cumulative Coulomb stress transfer. This ensemble of information allows us to place the sequence within the context of regional seismicity and to identify areas of remaining and/or elevated hazard. Our results constrain the size and spatial extent of rupture, and indicate that this was not the earthquake that had been anticipated. Significant sections of the northern Chile subduction zone have not ruptured in almost 150 years, so it is likely that future megathrust earthquakes will occur to the south and potentially to the north of the 2014 Iquique sequence.**

On 1 April 2014, a  $M 8.2$  earthquake ruptured a portion of the subduction zone in northern Chile offshore of the city of Iquique, a major port and hub for the country's copper mining industry. Peak shaking intensities reached MMI VIII on land, and a tsunami  $\sim 2$  m high hit coastal towns in southern Peru and northern Chile. Six fatalities were attributed to the event, and at least 13,000 homes were damaged or destroyed. Preliminary estimates suggest total economic losses close to US\$100 million<sup>7</sup>.

A megathrust earthquake in this region was not unexpected; 230  $M > 3.5$  earthquakes occurred offshore of Iquique between August 2013 and March 2014, a 950% increase in the rate from January to July 2013 (ref. 8). Over the three weeks before the event, there were more than 80 earthquakes between  $M 4.0$  and  $M 6.7$  (Fig. 1). Before the recent sequence, this subduction zone (between  $\sim 19.5^\circ$  S and  $21^\circ$  S) had been identified as a seismic gap<sup>3,4</sup>, last rupturing in a  $M \sim 8.8$  earthquake in 1877. The 1 April event was followed by a vigorous aftershock sequence with more than 100  $M \geq 4$  earthquakes, including a  $M 7.7$  aftershock near the southernmost extent of the  $M 8.2$  rupture.

The seismic moment of all 2014 earthquakes to date equates to an event of just  $M \sim 8.3$ , much smaller than the estimated size of the 1877 earthquake and of the potential event that could fill the seismic gap<sup>5,6</sup> (Methods). The earthquake sequence spans a section of the subduction zone about one-third of the size of the inferred 1877 rupture<sup>9</sup>. It remains unknown how subduction zones behave over multiple seismic cycles and whether any given section can be associated with a characteristic earthquake, making it unclear whether this seismic gap should behave in the twenty-first century as it did in the nineteenth. Observations suggest that enough strain has accumulated along this plate boundary segment to host an earthquake close to  $M 9$  (see, for example, ref. 5), and earthquakes of this size have occurred in the past. The expectation from

a seismic hazard perspective is that the fault can host another event of a similar magnitude. Although a great-sized earthquake here had been expected, it is possible that this event was not it<sup>10</sup>.

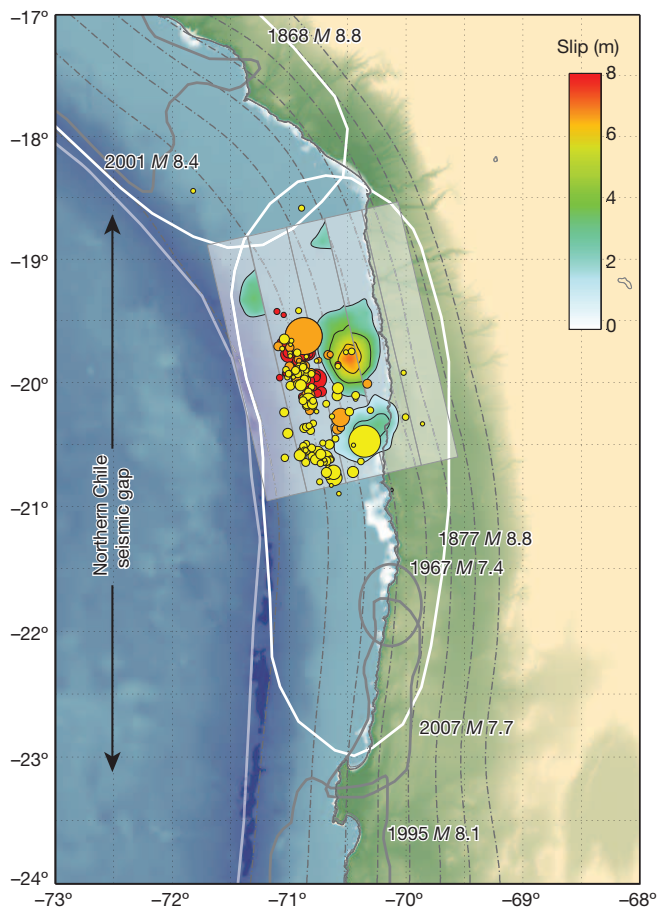
Sections of this subduction zone have ruptured since 1877 (Fig. 1), most notably in 1967, in a  $M 7.4$  event between  $\sim 21.5^\circ$  S and  $22^\circ$  S, and in the 2007  $M 7.7$  Tocopilla earthquake between  $\sim 22^\circ$  S and  $23.5^\circ$  S. Slip during these events was limited to the deeper extent of the seismogenic zone, leaving shallower regions unruptured<sup>6,11</sup>. Farther south, the 1995  $M 8.1$  Antofagasta earthquake broke the seismogenic zone immediately south of the Mejillones Peninsula, a feature argued to be a persistent barrier to rupture propagation<sup>11,12</sup>. Adjacent to the southern coast of Peru, a seismic gap associated with the 1868  $M 8.8$  rupture was partly filled by the 2001  $M 8.4$  Arequipa earthquake. Coupling models<sup>5</sup> indicate that strain accumulation may remain to the southeast of the Arequipa event, towards the northernmost edge of the 2014 rupture in Chile (a section  $\sim 200$  km long). However, within that zone adjacent to Arica, at the pronounced bend in the subduction zone and Peru–Chile Trench, coupling is low and may not support throughgoing rupture.

The National Earthquake Information Center (NEIC) and Centro Sismológico Nacional (CSN) W-phase centroid moment tensor (CMT) solutions for the 1 April 2014 earthquake align with the slab interface<sup>13</sup> and indicate a seismic moment of  $(1.00\text{--}2.35) \times 10^{21}$  N m ( $M_w = 8.07\text{--}8.18$ ). Our finite fault solution<sup>14</sup> (Methods and Fig. 2) describes a rupture area in the deeper portion of the seismogenic zone, with a peak slip of  $\sim 8$  m to the southeast of the hypocentre at depths of  $\sim 30\text{--}40$  km. Shallower slip to the north is not well resolved but may account for the generation of a local tsunami. Slip extended only  $\sim 50$  km along the interface from the hypocentre, a very compact rupture area for an earthquake of this size<sup>15</sup>. The location of peak slip in this model is consistent with W-phase CMT inversions (Extended Data Fig. 1), slab geometry (Extended Data Fig. 2) and the centroid location of an updated CMT (Methods).

This model also matches inversions of regional geodetic data (Methods and Extended Data Fig. 3), which do not uniquely resolve slip up-dip of the hypocentre but place strong constraints on the location and extent of slip between the hypocentre and the coast, and on its down-dip edge. Geodetic models show that slip during the mainshock ended west of the coastline, in agreement with (perhaps slightly up-dip of) seismic models. Tsunami models<sup>16</sup> place better constraint on shallow slip, and show little motion up-dip and west of the hypocentre.

Earthquakes in this sequence were relocated by using a multiple-event, hypocentroidal decomposition<sup>17</sup> algorithm, using seismic phase data from local, regional and global stations. This allows us to interpret locations within a consistent, regionally anchored and absolute framework<sup>2,18</sup>. Beginning on 16 March with a  $M 6.7$  earthquake, the foreshock sequence generated more than 80  $M \geq 4$  earthquakes (Fig. 2), showing a northward migration towards the epicentre of the 1 April  $M 8.2$  event (Extended Data Fig. 4). Over the following week, the NEIC recorded 140  $M \geq 4$  aftershocks, including a  $M 7.7$  event on 3 April, 27 h after the mainshock.

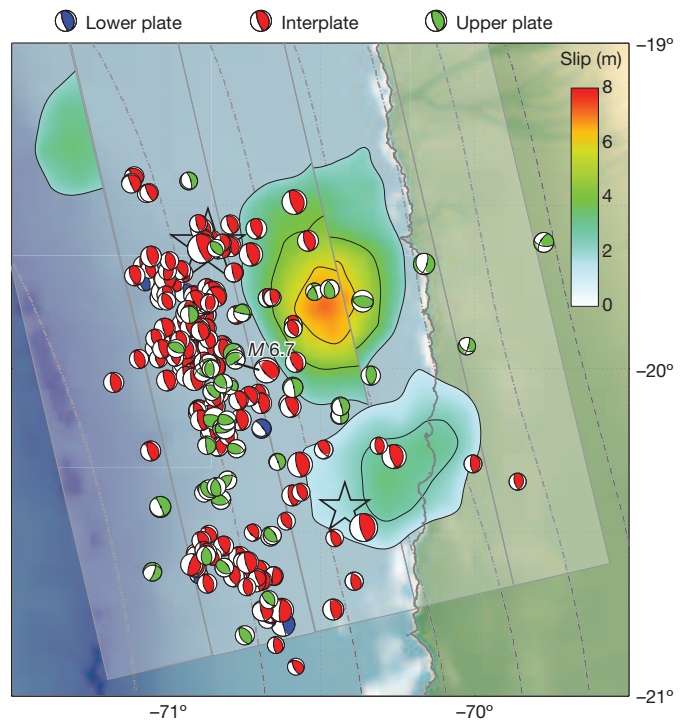
<sup>1</sup>National Earthquake Information Center, United States Geological Survey, Golden, Colorado 80401, USA. <sup>2</sup>Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>3</sup>Centro Sismológico Nacional, Universidad de Chile, Blanco Encalada 2002, Santiago 8370449, Chile. <sup>4</sup>Global Seismological Services, Golden, Colorado 80401, USA. <sup>5</sup>Canada Centre for Mapping and Earth Observation, Natural Resources Canada, Ottawa, Ontario K1A0E4, Canada.



**Figure 1 | Tectonic setting of 2014 Iquique earthquake sequence.** Rupture areas of large historical earthquakes are indicated by grey (modelled) and white (estimated) outlines. Relocated 2014 earthquakes are shown by colour: foreshocks in red; those between the mainshock (largest orange circle) on 1 April 2014 and the largest aftershock ( $M 7.7$ ) on 3 April in orange; and more recent events in yellow. Rupture areas of the  $M 8.2$  and  $M 7.7$  events are coloured and contoured at 2.0-m intervals. The extent of the northern Chile seismic gap is indicated with arrows. Bathymetric data are taken from the GEBCO\_08 grid<sup>30</sup>.

Between the mainshock and this largest aftershock, earthquake locations migrated southwards, towards the epicentre of the  $M 7.7$  event. Since then, aftershocks have been scattered up-dip of the rupture areas of the two largest events.

Analyses of regional moment tensors<sup>19</sup> (RMTs; Methods) of foreshocks indicate that many represent thrust faulting on or near the plate interface (Fig. 2). However, about 20% have well-constrained depths too shallow to involve interface slip, have focal mechanisms inconsistent with thrust faulting, or have nodal planes severely rotated with respect to local slab structure. Although constraining the absolute depths of offshore earthquakes in subduction zones is difficult, the spread of solutions (which are all subject to similar uncertainties) and the systematic rotation of many mechanisms with respect to slab geometry indicate that upper plate faulting was involved in the foreshock sequence. This is particularly true of the  $M 6.7$  event on 16 March, whose depth (20 km, NEIC; 20.6 km, CSN; 12 km, global centroid moment tensor project (global CMT)<sup>20</sup>; 15.5 km, NEIC W-phase) is substantially shallower than the slab, and whose shallow nodal plane is rotated  $60^\circ$  (global CMT; CSN W-phase) to  $70^\circ$  (NEIC W-phase) anticlockwise with respect to the slab, perhaps indicating that this foreshock occurred above the plate interface along a splay fault. Aftershock RMTs show that the vast majority occurred along the megathrust interface (Fig. 2), surrounding regions of largest co-seismic slip. The  $M 7.7$  aftershock ruptured a compact portion of the seismogenic zone  $\sim 50$  km south and directly along strike



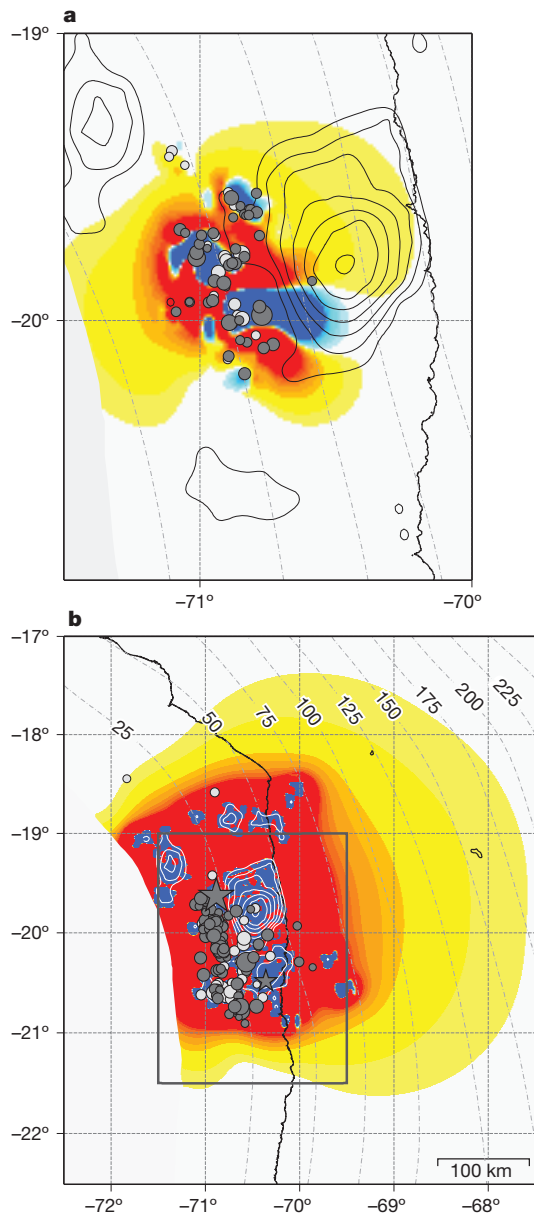
**Figure 2 | Source processes of events in the March–April 2014 Iquique earthquake sequence.** RMTs of relocated earthquakes in this sequence are shown and coloured by their location with respect to the slab interface; those interpreted as upper plate events are green, lower plate earthquakes are blue, and interplate events are red. Earthquakes are overlain on the preferred fault-slip models for the  $M 8.2$  and  $M 7.7$  events (hypocentres are shown with stars), with 2-m contour intervals. Dot-dashed lines in the background are slab contours<sup>13</sup>, plotted every 10 km.

from the mainshock asperity. Almost all 2014 events have been located up-dip of the rupture zones of the two biggest quakes.

Recent megathrust earthquake sequences (see, for example, refs 21, 22) have demonstrated the need for integrative real-time monitoring and assessments that map seismic cycles into models of strain accumulation near the source regions of large earthquakes. The Iquique sequence is an ideal case study involving the integration of geodetic, geodynamic and seismological constraints to improve the quantification and assessment of an earthquake sequence as it evolves. Organizing seismotectonic information for major global plate boundaries<sup>2,23</sup> is crucial for understanding the spectrum of expected behaviour of a fault zone after a major event has occurred (and indeed beforehand), and is the foundation of any framework for better communication of time-dependent earthquake hazards to affected communities<sup>24</sup>.

To understand the scale of this earthquake in the context of the Chilean subduction zone, it is useful to compare the Iquique sequence with the 2010  $M 8.8$  Maule earthquake<sup>2</sup>, the Chilean margin's last great megathrust event. Although both occurred in recognized seismic gaps, their evolution, behaviour and characteristics were quite different (Extended Data Fig. 4). In contrast to Iquique, Maule did not have any recognized foreshocks. Its mainshock nucleated in the middle of the South Central Chile Seismic Gap<sup>1</sup> and ruptured bilaterally beyond the extent of the gap into the 1985  $M 8.0$  and 1960  $M 9.5$  rupture zones to the north and south, respectively. The Iquique earthquake nucleated within the Northern Chile Seismic Gap, but in a region that had slipped more recently than much of the gap, and it ruptured an area much shorter than the gap's recognized extent<sup>5,6,11,25</sup>. Over the two months surrounding the Maule mainshock, earthquakes demonstrated a typical Gutenberg–Richter relationship, with a  $b$  value of  $\sim 0.85$  (Extended Data Fig. 4). The Iquique sequence, on the other hand, is deficient in moderate-to-large events ( $b \approx 0.73$ , in contrast with  $b \approx 1.02$  over the preceding year). Co-seismic<sup>2</sup> and post-seismic<sup>26</sup>



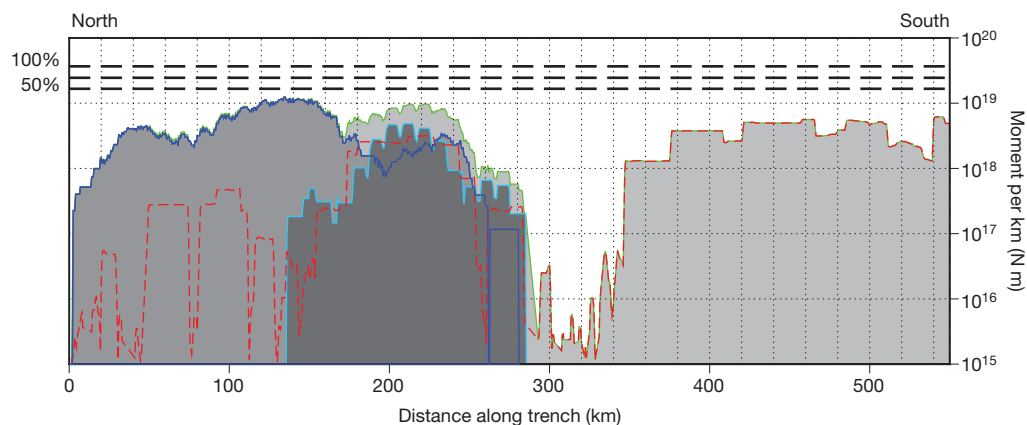


**Figure 3 | Coulomb failure ( $\Delta$ CFS) stress changes for foreshocks and aftershocks.**  $\Delta$ CFS is resolved onto the subduction zone interface<sup>13</sup> 1 day before the mainshock (a) and after the largest aftershock (b). Earthquakes are coloured by the local  $\Delta$ CFS resulting from previous earthquakes, at their time of occurrence; light grey symbols indicate negative stress changes, and dark grey symbols, positive. The slip models of the mainshock and the  $M$  7.7 aftershock are shown with white contours (1 m intervals). The dashed box in b represents the spatial extent of the region in a.

slip during the Maule sequence indicated rupture of most of the seismogenic zone, resulting in aftershock mechanisms that spanned a broad range of the faulting spectrum (Extended Data Fig. 4). In contrast, the Iquique sequence has not elicited a clear upper plate or outer rise response. Future studies of regional GPS data<sup>27</sup> may reveal the extent of post-seismic interface slip and its relation to the 2014 Iquique sequence.

We can use Coulomb failure stress ( $\Delta$ CFS) analysis<sup>28</sup> to assess whether the 2014 Iquique earthquake sequence followed a spatial and temporal migration pattern dictated by the stress changes caused by previous earthquakes.  $\Delta$ CFS calculations (Fig. 3) show that 18 of the 20 foreshocks with associated RMTs ruptured the megathrust interface where it had been positively stressed by previous events, and loaded the hypocentral region of the subsequent  $M$  8.2 event by 0.04 MPa. It thus seems that the northward migration of foreshocks responded to cascading  $\Delta$ CFS, ultimately leading to the mainshock. The hypocentral region of the  $M$  7.7 aftershock was loaded 0.25 MPa by the mainshock and the first 27 h of aftershocks. Aftershocks have generally nucleated in areas of increased  $\Delta$ CFS, surrounding the main slip patches of the largest events. Overall, the hypocentres of  $\sim 70\%$  of relocated aftershocks (94 of 138 events) occurred in areas of positive  $\Delta$ CFS. If uncertainties in relocated hypocentres ( $\pm 2$ – $3$  km) are considered (Fig. 3), more than 80% of aftershocks occurred in regions of positive  $\Delta$ CFS, lending support to the real-time use of  $\Delta$ CFS modelling as earthquake sequences unfold to aid in anticipating likely locations for subsequent events.

Analysis of moment accumulation and release along the plate margin (Fig. 4) shows that the main asperity ruptured in the  $M$  8.2 earthquake partly filled a historical gap in moment relative to adjacent sections of the arc. The March 2014 foreshocks began near the northern extent of the 1933 event, and moved northwards. Despite surrounding regions of larger moment deficit, the southern portion of the  $M$  8.2 and the  $M$  7.7 aftershock ruptured approximately the same region as the 1933 event. Moment from both large 2014 earthquakes, and from all historical events, falls off rapidly to the south between  $20^\circ$  S and  $21^\circ$  S; here, a 50–80-km section of arc has seen little to no seismic activity over the past century or more, until the northern extent of the 1967  $M$  7.4 earthquake is reached.



**Figure 4 | Moment deficit along the northern Chile subduction zone.** Moment calculated for historical seismicity from the USGS Combined Catalog<sup>32</sup> since 1900, resolved as moment per kilometre along strike. For each earthquake, moment is divided evenly over the length of the rupture, calculated using empirical relations<sup>14</sup>. For the largest earthquakes ( $M > 7.5$ ), more

accurate rupture areas are used<sup>8</sup>. Red shows moment for historical earthquakes; blue for 2014 seismicity (dark blue up to 3 April 2014; light blue since then); green represents all summed moment. Horizontal dashed lines represent moment accumulation levels given constant coupling percentages of 50%, 75% and 100%.

Continuing south, parts of the seismogenic zone have been ruptured by this 1967 event and the 2007  $M_{w}$  7.7 Tocopilla earthquake, although much of the shallow portion of the subduction zone here has remained unruptured since at least 1877.

These analyses imply that there is cause for concern for one or more further megathrust earthquakes in northern Chile. Large earthquakes since 1877, in 1967 and 2007 in the south, and now with this 2014 sequence in the north, have combined to rupture a fraction of the Northern Chile Seismic Gap, only partly releasing accumulated strain (Fig. 4). The highly coupled section of the subduction zone  $\sim 50$ – $80$  km to the south of the 2014 events, and the moderately coupled interface for a similar distance to the north, have been positively stressed by the 2014 Iquique sequence (Fig. 3). Neither section has hosted significant events in almost 150 years. Given that earthquakes in this subduction zone have occurred in response to stress transfer in the past, there is now an increased probability of megathrust earthquakes occurring to the south or north of the April 2014 events in the future.

Comparisons between the 2014 Iquique and 2010 Maule sequences highlight the broad range of seismotectonic behaviour that is possible along the same subduction zone, leading up to and in response to megathrust ruptures. Until we understand these differences, our ability to accurately predict the future behaviour of surrounding megathrust regions<sup>29</sup> will remain elusive. Nevertheless, it is likely that the subduction zone to the south of the March–April 2014 events, between Iquique and Antofagasta, will host great-sized megathrust events in the future. A smaller section to the north, towards Arica, may also be capable of hosting a  $M \geq 8$  event. Chilean and global seismologists now face the difficult task of communicating this uncertain yet perhaps elevated hazard, without appearing alarmist.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 April; accepted 10 July 2014.**

**Published online 13 August 2014.**

- McCann, W. R., Nishenko, S. P., Sykes, L. R. & Krause, J. Seismic gaps and plate tectonics: seismic potential for major boundaries. *Pure Appl. Geophys.* **117**, 1082–1147 (1979).
- Hayes, G. P. *et al.* Seismotectonic framework of the February 27, 2010  $M_{w}$  8.8 Maule, Chile earthquake sequence. *Geophys. J. Int.* **105**, 1034–1051 (2013).
- Comte, D. & Pardo, M. Reappraisal of great historical earthquakes in the northern Chile and southern Peru seismic gaps. *Nat. Hazards* **4**, 23–44 (1991).
- Loveless, J. P. & Meade, B. J. Spatial correlation of interseismic coupling and coseismic rupture extent of the 2011  $M_{w}$  = 9.0 Tohoku-oki earthquake. *Geophys. Res. Lett.* **38**, L17306 (2011).
- Chlieh, M. *et al.* Interseismic coupling and seismic potential along the Central Andes subduction zone. *J. Geophys. Res.* **116**, B12405 (2011).
- Béjar-Pizarro, M. *et al.* Andean structural control on interseismic coupling in the North Chile subduction zone. *Nature Geosci.* **6**, 462–467 (2013).
- Podlaha, A. & Bowen, S. *Impact Forecasting: April 2014 Global Catastrophe Recap, AON Benfield Report* <[http://thoughtleadership.aonbenfield.com/Documents/20140507\\_if\\_april\\_global\\_recap.pdf](http://thoughtleadership.aonbenfield.com/Documents/20140507_if_april_global_recap.pdf)> (April 2014).
- Centro Sismológico Nacional, Universidad de Chile <<http://www.sismologia.cl/>>.
- Kausel, E. Los terremotos de Agosto de 1868 y Mayo de 1877 que afectaron el Sur del Perú y Norte de Chile. *Boll. Acad. Chil. Ciencias* **3**, 8–12 (1986).
- Witze, A. Chile quake defies expectations. *Nature* **508**, 440–441 (2014).
- Béjar-Pizarro, M. *et al.* Asperities and barriers on the seismogenic zone in the North Chile: state-of-the-art after the 2007  $M_{w}$  7.7 Tocopilla earthquake inferred by GPS and InSAR data. *Geophys. J. Int.* **183**, 390–406 (2010).
- Armijo, R. & Thiele, R. Active faulting in Northern Chile: ramp stacking and lateral decoupling along a subduction plate boundary? *Earth Planet. Sci. Lett.* **98**, 40–61 (1990).
- Hayes, G. P., Wald, D. J. & Johnson, R. L. Slab1.0: a three-dimensional model of global subduction zone geometries. *J. Geophys. Res.* **117**, B01302 (2012).
- Ji, C., Wald, D. J. & Helmberger, D. V. Source description of the 1999 Hector Mine, California, earthquake, part I: wavelet domain inversion theory and resolution analysis. *Bull. Seismol. Soc. Am.* **92**, 1192–1207 (2002).
- Blaser, L., Krüger, F., Ohrnberger, M. & Scherbaum, F. Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bull. Seismol. Soc. Am.* **100**, 2914–2926 (2010).
- Lay, T., Yue, H., Brodsky, E. E. & An, C. The 1 April, 2014 Iquique, Chile  $M_{w}$  8.1 earthquake rupture sequence. *Geophys. Res. Lett.* **41**, 3818–3825 (2014).
- Jordan, T. H. & Sverdrup, K. A. Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific. *Bull. Seismol. Soc. Am.* **71**, 1105–1130 (1981).
- Walker, R. T., Bergman, E., Jackson, J., Ghorashi, M. & Talebian, M. The 2002 June 22 Changuh (Avaj) earthquake in Qazvin province, northwest Iran: epicentral relocation, source parameters, surface deformation and geomorphology. *Geophys. J. Int.* **160**, 707–720 (2005).
- Herrmann, R. B., Benz, H. & Ammon, C. J. Monitoring the earthquake source process in North America. *Bull. Seismol. Soc. Am.* **101**, 2609–2625 (2011).
- Ekström, G., Nettles, M. & Dziewoński, A. M. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Inter.* **200**, 1–9 (2012).
- Lay, T. *et al.* The 2006–2007 Kuril Islands great earthquake sequence. *J. Geophys. Res.* **114**, B11308 (2009).
- Hayes, G. P., Furlong, K. P., Benz, H. M. & Herman, M. W. Triggered aseismic slip adjacent to the 6 February 2013  $M_{w}$  8.0 Santa Cruz Islands megathrust earthquake. *Earth Planet. Sci. Lett.* **388**, 265–272 (2014).
- Hayes, G. P. *et al.* 88 hours: the U.S. Geological Survey National Earthquake Information Center response to the 11 March 2011  $M_{w}$  9.0 Tohoku earthquake. *Seismol. Res. Lett.* **82**, 481–493 (2011).
- Jordan, T. H. & Jones, L. M. Operational earthquake forecasting: some thoughts on why and how. *Seismol. Res. Lett.* **81**, 571–574 (2010).
- Metois, M. *et al.* Revisiting the North Chile seismic gap segmentation using GPS-derived interseismic coupling. *Geophys. J. Int.* **194**, 1283–1294 (2013).
- Bedford, J. *et al.* A high-resolution, time-variable afterslip model for the 2010 Maule  $M_{w}$  = 8.8, Chile megathrust earthquake. *Earth Planet. Sci. Lett.* **383**, 26–36 (2013).
- Integrated Plate Boundary Observatory Chile. <<http://www.ipoc-network.org/>>.
- Stein, R. S., Barka, A. A. & Dieterich, J. H. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **128**, 594–604 (1997).
- Brodsky, E. E. & Lay, T. Recognizing foreshocks from the April 2014 Chile earthquake. *Science* **344**, 700–702 (2014).
- GEBCO. *General Bathymetric Chart of the Oceans* <<http://www.gebco.net>>.
- Wessel, P. & Smith, W. H. F. Free software helps map and display data. *EOS* **72**, 445–446 (1991).
- USGS. *About ANSS Comprehensive Catalog* <[http://earthquake.usgs.gov/earthquakes/map/doc\\_aboutdata.php](http://earthquake.usgs.gov/earthquakes/map/doc_aboutdata.php)>.

**Acknowledgements** We thank R. Briggs for his comments in improving the manuscript. We thank the CSN, and member institutions of the IPOC network for their operation of seismic stations in northern Chile and for the contribution of waveform data and phase picks to this study. This study made use of broadband seismic data from globally distributed seismometers available to the USGS NEIC in real time or near real time (networks AE, BK, C, CN, CU, DK, G, GE, GT, IC, II, IU, IW and US) and archived in the NEIC Central Waveform Buffer and at the Incorporated Research Institutions for Seismology Data Management Center. RADARSAT-2 data were provided by the Canadian Space Agency and MDA Corporation. Bathymetry data come from GEBCO 2008 (ref. 30). Many of the figures were made with the Generic Mapping Tools software package<sup>31</sup>. National Science Foundation grant EAR-1153317 provided support to K.P.F. and M.W.H. for this research. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

**Author Contributions** G.P.H., K.P.F. and W.D.B. wrote the manuscript. G.P.H. generated Figs 1, 2 and 4; M.W.H. generated Fig. 3. G.P.H., M.W.H. and W.D.B. generated Extended Data figures. G.P.H. conducted seismic fault inversions and moment deficit calculations. M.W.H. conducted stress transfer calculations. W.D.B. conducted geodetic fault inversions. G.P.H. and W.D.B. were jointly responsible for fault models. S.R. and S.B. contributed regional real-time analyses and data used in seismic inversions and earthquake relocations. E.B. and H.M.B. conducted earthquake relocations. P.S.E. contributed to the real-time analysis of the earthquake sequence and edited the manuscript. S.S. scheduled the acquisition of RADARSAT-2 data and performed InSAR analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.P.H. (ghayes@usgs.gov).



# Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake

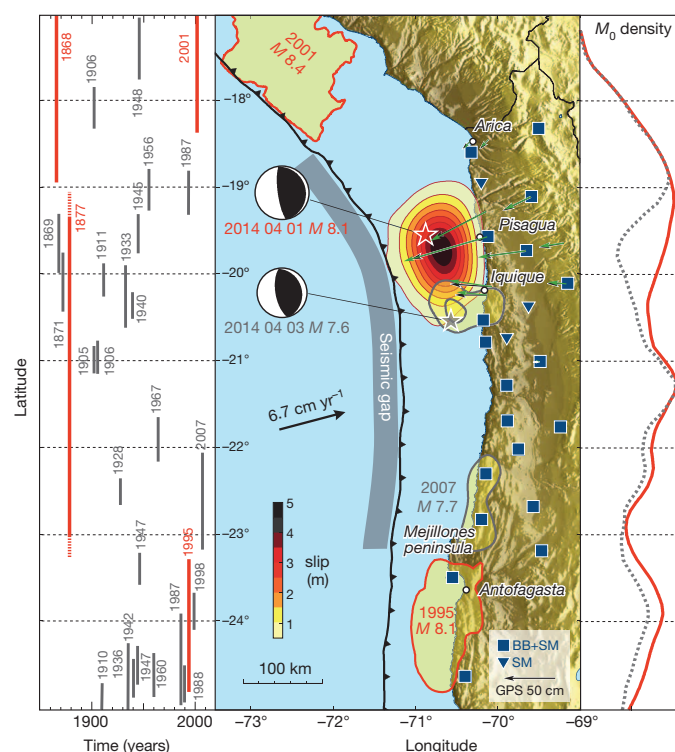
Bernd Schurr<sup>1</sup>, Günter Asch<sup>1</sup>, Sebastian Hainzl<sup>1</sup>, Jonathan Bedford<sup>1</sup>, Andreas Hoechner<sup>1</sup>, Mauro Palo<sup>1</sup>, Rongjiang Wang<sup>1</sup>, Marcos Moreno<sup>1</sup>, Mitja Bartsch<sup>1</sup>, Yong Zhang<sup>2</sup>, Onno Oncken<sup>1</sup>, Frederik Tilmann<sup>1</sup>, Torsten Dahm<sup>1</sup>, Pia Victor<sup>1</sup>, Sergio Barrientos<sup>3</sup> & Jean-Pierre Vilotte<sup>4</sup>

On 1 April 2014, Northern Chile was struck by a magnitude 8.1 earthquake following a protracted series of foreshocks. The Integrated Plate Boundary Observatory Chile monitored the entire sequence of events, providing unprecedented resolution of the build-up to the main event and its rupture evolution. Here we show that the Iquique earthquake broke a central fraction of the so-called northern Chile seismic gap, the last major segment of the South American plate boundary that had not ruptured in the past century<sup>1,2</sup>. Since July 2013 three seismic clusters, each lasting a few weeks, hit this part of the plate boundary with earthquakes of increasing peak magnitudes. Starting with the second cluster, geodetic observations show surface displacements that can be associated with slip on the plate interface. These seismic clusters and their slip transients occupied a part of the plate interface that was transitional between a fully locked and a creeping portion. Leading up to this earthquake, the *b* value of the foreshocks gradually decreased during the years before the earthquake, reversing its trend a few days before the Iquique earthquake. The mainshock finally nucleated at the northern end of the foreshock area, which skirted a locked patch, and ruptured mainly downdip towards higher locking. Peak slip was attained immediately downdip of the foreshock region and at the margin of the locked patch. We conclude that gradual weakening of the central part of the seismic gap accentuated by the foreshock activity in a zone of intermediate seismic coupling was instrumental in causing final failure, distinguishing the Iquique earthquake from most great earthquakes. Finally, only one-third of the gap was broken and the remaining locked segments now pose a significant, increased seismic hazard with the potential to host an earthquake with a magnitude of >8.5.

The northern Chile–southern Peru seismic gap last broke in 1877 in a great earthquake ( $M_w \sim 8.8$ )<sup>1</sup> that ruptured from south of Arica to the Mejillones peninsula (see Fig. 1). The reported historical recurrence interval for the past 500 years in this region has been estimated at  $111 \pm 33$  years<sup>1</sup>, making it probably the most mature seismic gap along the South American plate boundary. In the past two decades the two adjoining segments south and north broke in the  $M_w$  8.1 Antofagasta earthquake of 1995 (refs 3, 4) and the  $M_w$  8.4 Arequipa earthquake of 2001 in southern Peru<sup>5</sup>. In the previous cycle the southern Peru and northern Chile segments broke within few years (Fig. 1), suggesting that they might be coupled in time<sup>1</sup>. The imminence of a large megathrust event in this region motivated the setting up of an international monitoring effort with the Integrated Plate Boundary Observatory Chile (IPOC). Having started in 2007, there now exists an exceptional database that monitors the gradual plate boundary failure with various geophysical techniques.

Several major earthquakes ( $M_w > 7$ ) have occurred in this gap since 1850 (Fig. 1); the largest until now was the  $M_w$  7.7 Tocopilla earthquake in 2007, which broke the southern rim of this segment beneath and north of Mejillones peninsula along a total length of 150 km. Only the downdip end of the locked zone slipped in this event, and the total slip in the

rupture area was less than 2.6 m (refs 6, 7) leaving most of the past slip deficit of  $\sim 8$ –9 m accumulated since 1877 unaffected<sup>8</sup>. On 1 April 2014, the  $M_w$  8.1 Iquique earthquake north of Iquique struck the central portion of the gap. Using seismological and geodetic observations we here analyse the rupture, its relationship to previous locking of the plate interface, and the pre-seismic transients leading up to the earthquake.



**Figure 1 | Map of Northern Chile and Southern Peru showing historical earthquakes and instrumentally recorded megathrust ruptures.** IPOC instruments used in the present study (BB, broadband; SM, strong motion) are shown as blue symbols. Left: historical<sup>1,2</sup> and instrumental earthquake record. Centre: rupture length was calculated using the regression suggested in ref. 28, with grey lines for earthquakes  $M > 7$  and red lines for  $M_w > 8$ . The slip distribution of the 2014 Iquique event and its largest aftershock derived in this study are colour coded, with contour intervals of 0.5 m. The green and black vectors are the observed and modelled horizontal surface displacements of the mainshock. The slip areas of the most recent other large ruptures<sup>4,5,7</sup> are also plotted. Right: moment deficit per kilometre along strike left along the plate boundary after the Iquique event for moment accumulated since 1877, assuming current locking (Fig. 3a). The total accumulated moment since 1877 from 17° S to 25° S (red solid line) is 8.97; the remaining moment after subtracting all earthquake events with  $M_w > 7$  (grey dotted line) is 8.91 for the entire northern Chile–southern Peru seismic gap.

<sup>1</sup>GFZ Helmholtz Centre Potsdam, German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany. <sup>2</sup>School of Earth and Space Sciences, Peking University, Beijing 100871, China. <sup>3</sup>Centro Sismológico Nacional, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Blanco Encalada 2002, Santiago, Chile. <sup>4</sup>Institut de Physique du Globe de Paris, 1, rue Jussieu, 75238 Paris cedex 05, France.

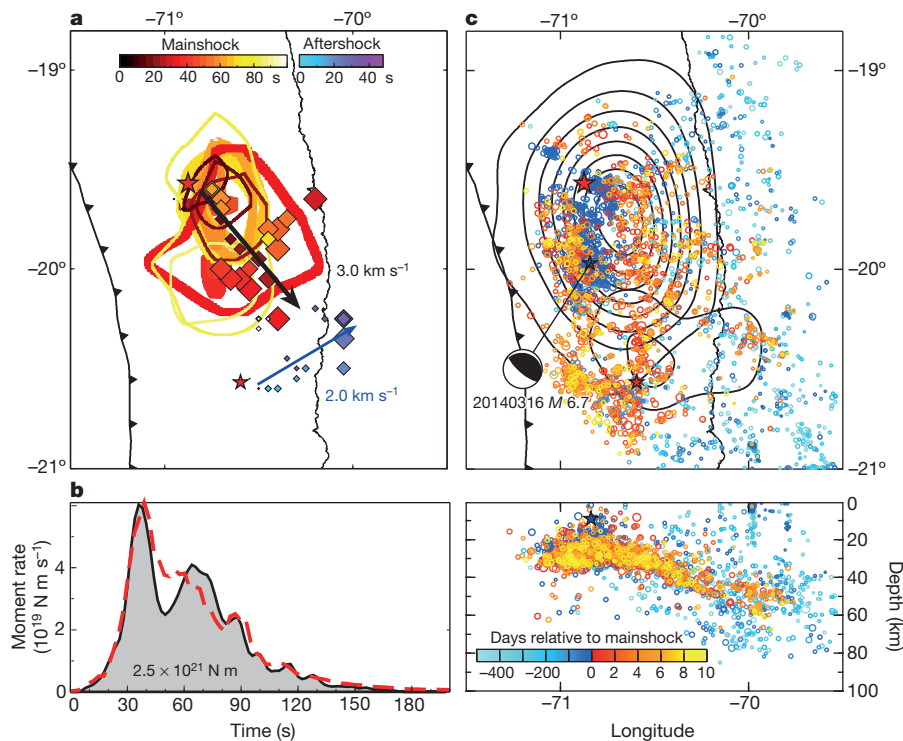
Kinematic analysis of the  $M_w$  8.1 Iquique rupture makes use of two complementary approaches. First, we performed waveform modelling of local strong motion seismograms and teleseismic body waves to constrain the kinematic development of the rupture towards the final displacements in a joint inversion with continuous GPS data of static displacements (Figs 1 and 2a). Second, we used the backprojection technique applied to stations in North America to map the radiation of high-frequency seismic waves (HFSR; 1–4 Hz)<sup>9,10</sup>. The latter technique is not sensitive to absolute slip amplitudes, but rather to changes in slip and rupture velocity. During the first 35–40 s the rupture propagated downdip with increasing velocity, nearly reaching the coastline (Fig. 2a, b). Towards the end of the rupture, the area near the epicentre was reactivated. In spite of the relatively complicated kinematic history of the rupture, the cumulative slip shows a simple ‘bullseye’ pattern with a peak co-seismic slip of about 4.4 m (Fig. 1), consistent with the slip model in ref. 11 based on teleseismic and deep-water tsunami wave recordings. A slip patch of similar size and magnitude has been found<sup>12</sup> but with the highest slip placed ~40 km further south from teleseismic-only recordings; however, this seems not to be compatible with onshore displacement measurements from GPS. The largest aftershock so far was the  $M_w$  7.6 event of 3 April that, like the mainshock, ruptured initially downdip. Rupture then propagated northeastwards, attaining a peak slip of 1.2 m after about 20 s. Both earthquakes broke a total of 200 km of the margin.

The Iquique rupture affected an area shown from geodetic analysis to be a zone of intermediate interseismic coupling at 18.5°–21° S interrupting the otherwise fully locked northern Chile–southern Peru gap<sup>8,13</sup> (Fig. 3a). The Iquique mainshock nucleated at the northwestern border of a locked patch and ruptured towards its centre (Figs 2a and 3a). The downdip end of the mainshock and for the large  $M_w$  7.6 aftershock rupture

mapped by both the HFSR and co-seismic slip agrees quite accurately with the downdip end interseismic coupling (Figs 2a, c and 3a). The accelerated downdip rupture propagation for both earthquakes closely followed the gradient towards higher locking. The Iquique event and its largest aftershock therefore seem to have broken the central, only partly locked segment of the northern Chile–southern Peru seismic gap, releasing part of the slip deficit accumulated here since 1877 (compare Fig. 1).

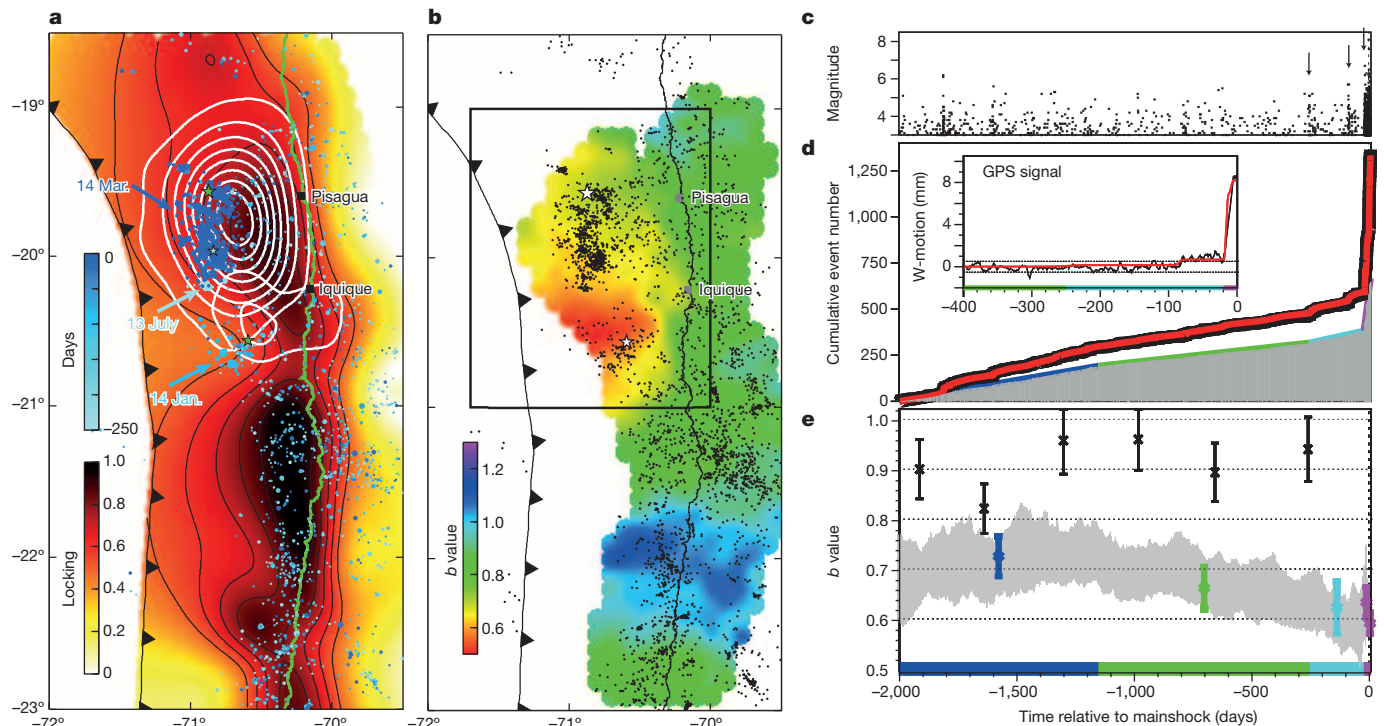
The seismicity before the Iquique earthquake also concentrated in this zone of intermediate locking at the fringe of the highly locked, high-slip patch (Fig. 3a). Starting in July 2013, three foreshock clusters with increasingly larger peak magnitudes and cumulative seismic moment occurred here (Figs 2c and 3a, c). The mainshock rupture started at the northern end of the foreshock zone, inside the region of intermediate locking (Figs 2c and 3a). The second foreshock cluster (January 2014) is associated with a weak transient deformation, whereas the third cluster (March 2014) shows a very distinct transient signal. GPS displacement vectors calculated over the times spanning these foreshock clusters point towards the cluster epicentres (Extended Data Fig. 4). Deformation for both transients is entirely explained by the cumulative co-seismic displacement of the respective foreshock clusters (Fig. 3d inset and Extended Data Fig. 4). The area affected by the foreshocks is then reactivated by the aftershock sequence of the Iquique earthquake (Fig. 2c). Comparing this with the long-term deformation history of the margin, we found that the same area shows a high gradient of locking from weakly locked updip to fully locked downdip (Fig. 3a).

Additionally, the analysis of the frequency–magnitude distribution of the foreshocks reveals outstanding characteristics in both the spatial and temporal vicinities of the rupture. The earthquake number  $N$  as a function of the magnitude  $M$  is found to follow the well-known Gutenberg–Richter



**Figure 2 | Kinematic rupture development of the  $M_w$  8.1 main and  $M_w$  7.6 aftershock and the distribution of foreshocks and aftershocks.** The nucleation point of each earthquake rupture is indicated by a coloured star. **a**, Arrows indicate the propagation of main energy release during the first 40 and 25 s for mainshock and aftershock, respectively. The contour lines represent isolines of slip rate for the mainshock from the kinematic inversion during different time intervals after rupture nucleation (0.05 m s<sup>-1</sup> intervals, line thickness scaled by slip-rate). Coloured diamonds represent maxima of semblance scaled to the peak value of the emitted energy for mainshock and aftershock for each time step based on the backprojection of teleseismic

waveforms. **b**, Moment rate and time history of backprojected energy (arbitrary absolute scale). Black solid line, mainshock kinematic source-time-function; red dashed line, rescaled backprojection energy. **c**, Map (top) and longitudinal cross-section (bottom) of ~3,600 foreshocks and ~1,400 aftershocks coloured according to their time of occurrence relative to the mainshock. The slips of the mainshock and largest aftershock are contoured. The beachball depicts the double-couple of the largest  $M_w$  6.7 foreshock that had a rupture geometry distinctively different from the mainshock and largest aftershock (Fig. 1) and a centroid depth of only 9 km (blue star) and that thus probably occurred in the upper plate.



**Figure 3 | Maps of interseismic locking and  $b$  value, and time history of seismicity and deformation.** **a**, Geodetic interseismic locking and foreshocks. The July–August 2013, January and March 2014 foreshock clusters are marked. **b**,  $b$ -value map of the central portion of northern Chile gap for the last 2000 days before the mainshock, where results are calculated for all  $M \geq 3$  foreshocks within 50 km if their number exceeds 100. The rectangular box encloses the area used for the results in **c–e**. **c**, Magnitude–time plot. Arrows mark the July–August 2013, January and March 2014 clusters. **d**, Observed (black thick line) and ETAS-modelled (red line) cumulative  $M \geq 3$  activity; the thin coloured

lines are fits of the estimated background (grey shaded area) for the four phases, during each of which the background rate is almost constant (see the text). The inset shows measured GPS displacement time-series stacked from near-coast stations between  $19^\circ$  S and  $21^\circ$  S smoothed with a four-day moving average and the modelled signal related to cumulative slip of the foreshock events. **e**, Time series of  $b$  value (means  $\pm$  s.d.) for the events inside the box (coloured and grey shaded area). Black points and bars refer to the results for the events outside the box.

relation,  $N \approx 10^{-bM}$ .  $b$  values have been proposed to act as a stress sensor, with low  $b$  values indicating high stresses<sup>14</sup>. Mapping the  $b$  value in Fig. 3b indicates significantly lower  $b$  values in the source area than in all other regions where the  $b$  value can be resolved. A gradual decrease in the  $b$  value from about 0.75 to below 0.6 is observed in the source region within the three years before the Iquique earthquake (Fig. 3e). This only reverses within the last days of the foreshock sequence. Similar decreases in  $b$  values before large megathrust earthquakes have recently been documented, in particular for the  $M$  9 Tohoku event<sup>15</sup>.

To reveal potential changes of background seismic activity related to aseismic processes, we fit the foreshock seismicity by means of an Epidemic-Type-Aftershock-Sequence (ETAS) model<sup>16</sup>, identifying only 42% of pre-mainshock events as Omori-type aftershocks triggered by larger foreshocks. Using the Akaike information criterion<sup>17</sup>, the remaining background rate is found to be significantly time dependent: we identified four subsequent periods of almost constant background rates (Fig. 3d, e). The mainshock preparation process seems to have been initiated by a relative seismic quiescence, which starts at the same time as  $b$  values are found to drop and ends in July 2013, when the background rate returns to approximately the pre-quiescence values; the final phase starts 18 days before the quake, when the background seismicity increases more than 35-fold in conjunction with the onset of the final transient GPS signal.

Such a sequence of seismic quiescence, recovery and acceleration of background activity is expected in the stress accumulation framework<sup>18</sup>. On the basis of the inverted ETAS parameters and the low  $b$  values before the mainshock, the seismicity is expected to accelerate with time as the result of a branching ratio—that is, the average number of daughter events per earthquake—larger than one indicating a transient supercritical state<sup>19</sup>. Hence, the period leading to the Iquique earthquake documents progressive asperity failure, here observed in unprecedented detail. An increase

in foreshock activity preceding subduction megathrust events has repeatedly been reported<sup>18,20,21</sup>. In contrast to, for example, the foreshocks of the 2011  $M$  9 Tohoku–Oki earthquake<sup>21</sup>, which started about three weeks before the mainshock, the foreshock clusters described here have been active since the start of our observations in 2007. However, only more recently have foreshocks with increasing magnitudes and thus more fertile aftershock sequences resulted in a supercritical state with self-accelerating seismicity<sup>19</sup>. This ultimately allowed an earthquake to nucleate that was strong enough to break into the stronger locked and thus aseismic part of the interface. From the spatial correlation of foreshock activity and slip gradients with the Iquique mainshock rupture region, we infer gradual unlocking of the plate interface. Moreover, the start of the decrease in  $b$  value  $\sim 3$  years ago, in spite of constant plate convergence and loading rate, indicates a physical change at the plate interface. From only intermediate locking the Iquique segment seems to have been dominated by mostly smaller locked asperities embedded in a conditionally stable environment, if we assume that rate and state frictional behaviour is controlling locking and creep<sup>22–24</sup>. Progressive rupturing of the smaller asperities by foreshocks will have loaded the remaining larger asperities in this zone until their failure: this evolution may be seen as the culmination of a runaway process as the likely key mechanism leading up to the Iquique earthquake. Both the rupture direction and speed of the mainshock and its triggered large aftershock are controlled by the stress gradient in the remaining asperities, corroborating theoretical analysis<sup>25</sup>. This indicates that not only the size of larger asperities but also their stress topography is important for understanding the propagation, acceleration and stopping of megathrust earthquakes. The Iquique event broke a region of heterogeneous coupling, where rupture not only broke a moderate-sized asperity but also penetrated into a weakly coupled zone, possibly by dynamic weakening<sup>26</sup>.



Earthquake swarms occur preferentially in regions of low interseismic strain accumulation, which tend to delimit subsequent ruptures of plate interface zones<sup>27</sup>. The  $M_w$  7.7 Tocopilla earthquake with its aftershocks and the  $M_w$  8.1 Iquique earthquake with its foreshocks might be considered similar features. They embrace the rims of a highly locked zone 250 km long in the centre of the northern Chile–southern Peru gap with an accumulated slip deficit<sup>8</sup> of some 8–9 m since 1877. Estimating the moment deficit for the entire seismic gap yields the potential for an earthquake with a maximum magnitude of  $M_w \sim 8.9$  (Fig. 1) for the historically less likely case that the entire zone fails in a single event. However, the reduction of the slip deficit by about 50% in the Iquique earthquake area (Fig. 1) decreases the probability that a future earthquake will release the whole remaining slip deficit at once. As the slip deficit reduction is only partial, though, this region will not necessarily act as a barrier: the seismic potential of this area remains high.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 23 May; accepted 14 July 2014.**

**Published online 13 August 2014.**

- Comte, D. & Pardo, M. Reappraisal of great historical earthquakes in the northern Chile and southern Peru seismic gaps. *Nat. Hazards* **4**, 23–44 (1991).
- Chlieh, M. *et al.* Interseismic coupling and seismic potential along the Central Andes subduction zone. *J. Geophys. Res.* **116**, 1–21 (2011).
- Ruegg, J. C. *et al.* The  $M_w = 8.1$  Antofagasta (North Chile) Earthquake of July 30, 1995: first results from teleseismic and geodetic data. *Geophys. Res. Lett.* **23**, 917–920 (1996).
- Chlieh, M. *et al.* Crustal deformation and fault slip during the seismic cycle in the North Chile subduction zone, from GPS and InSAR observations. *Geophys. J. Int.* **158**, 695–711 (2004).
- Pritchard, M. E. *et al.* Geodetic, teleseismic, and strong motion constraints on slip from recent southern Peru subduction zone earthquakes. *J. Geophys. Res.* **112**, 1–24 (2007).
- Delouis, B., Pardo, M., Legrand, D. & Monfret, T. The  $M_w$  7.7 Tocopilla earthquake of 14 November 2007 at the southern edge of the northern Chile seismic gap: rupture in the deep part of the coupled plate interface. *Bull. Seismol. Soc. Am.* **99**, 87–94 (2009).
- Schurr, B. *et al.* The 2007  $M_w$  7.7 Tocopilla northern Chile earthquake sequence: Implications for along-strike and downdip rupture segmentation and megathrust frictional behavior. *J. Geophys. Res.* **117**, 1–19 (2012).
- Métois, M. *et al.* Revisiting the North Chile seismic gap segmentation using GPS-derived interseismic coupling. *Geophys. J. Int.* **194**, 1283–1294 (2013).
- Krüger, F. & Ohrnberger, M. Tracking the rupture of the  $M_w = 9.3$  Sumatra earthquake over 1,150 km at teleseismic distance. *Nature* **435**, 937–939 (2005).
- Rössler, D., Krüger, F., Ohrnberger, M. & Ehlert, L. Rapid characterisation of large earthquakes by multiple seismic broadband arrays. *Nat. Hazards Earth Syst. Sci.* **10**, 923–932 (2010).
- Lay, T., Yue, H., Brodsky, E. E. & An, C. The 1 April 2014 Iquique, Chile,  $M_w$  8.1 earthquake rupture sequence. *Geophys. Res. Lett.* **41**, 3818–3825 (2014).
- Yagi, Y. *et al.* Rupture process of the 2014 Iquique Chile Earthquake in relation with the foreshock activity. *Geophys. Res. Lett.* **41**, 1–6 (2014).
- Béjar-Pizarro, M. *et al.* Andean structural control on interseismic coupling in the North Chile subduction zone. *Nature Geosci.* **6**, 462–467 (2013).
- Schorlemmer, D., Wiemer, S. & Wyss, M. Variations in earthquake-size distribution across different stress regimes. *Nature* **437**, 539–542 (2005).
- Nanjo, K. Z., Hirata, N., Obara, K. & Kasahara, K. Decade-scale decrease in  $b$  value prior to the  $M_9$ -class 2011 Tohoku and 2004 Sumatra quakes. *Geophys. Res. Lett.* **39**, L20304 (2012).
- Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**, 9–27 (1988).
- Marsan, D., Prono, E. & Helmstetter, A. Monitoring aseismic forcing in fault zones using earthquake time series. *Bull. Seismol. Soc. Am.* **103**, 169–179 (2013).
- Mignan, A. Seismicity precursors to large earthquakes unified in a stress accumulation framework. *Geophys. Res. Lett.* **39**, L21308 (2012).
- Helmstetter, A., Kagan, Y. Y. & Jackson, D. D. Importance of small earthquakes for stress transfers and earthquake triggering. *J. Geophys. Res.* **110**, 1–13 (2005).
- Bouchon, M., Durand, V., Marsan, D., Karabulut, H. & Schmittbuhl, J. The long precursory phase of most large interplate earthquakes. *Nature Geosci.* **6**, 299–302 (2013).
- Kato, A. *et al.* Propagation of slow slip leading up to the 2011  $M_w$  9.0 Tohoku-Oki earthquake. *Science* **335**, 705–708 (2012).
- Scholz, C. H. Earthquakes and friction laws. *Nature* **391**, 37–42 (1998).
- Bilek, S. L. & Lay, T. Tsunami earthquakes possibly widespread manifestations of frictional conditional stability. *Geophys. Res. Lett.* **29**, 1–4 (2002).
- Kaneko, Y., Avouac, J. P. & Lapusta, N. Towards inferring earthquake patterns from geodetic observations of interseismic coupling. *Nature Geosci.* **3**, 363–369 (2010).
- Das, S. & Kostrov, B. V. Breaking of a single asperity: rupture process and seismic radiation. *J. Geophys. Res.* **88**, 4277–4288 (1983).
- Noda, H. & Lapusta, N. Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature* **493**, 518–521 (2013).
- Holtkamp, S. & Brudzinski, M. R. Megathrust earthquake swarms indicate frictional changes which delimit large earthquake ruptures. *Earth Planet. Sci. Lett.* **390**, 234–243 (2014).
- Dorbath, L., Cisternas, A. & Dorbath, C. Assessment of the size of large and great historical earthquakes in Peru. *Bull. Seismol. Soc. Am.* **80**, 551–576 (1990).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Data used in this study come from the IPOC initiative (<http://www.ipoc-network.org>) operated by the GFZ – German Research Centre for Geosciences, Institut de Physique du Globe de Paris, Centro Sismológico Nacional, Universidad de Chile, and Universidad Católica del Norte, Antofagasta, Chile. We also acknowledge the French–Chilean International Associated Laboratory (LIA) ‘Montessus de Ballore’ and the USA–Chilean Central Andean Tectonic Observatory Geodetic Array projects for giving access to data of several of their continuous GPS (cGPS) stations in Chile. Part of this work was made possible by the Hazard Assessment and Risk Team (HART) initiative funded by the GFZ and Hannover Re.

**Author Contributions** B.S. processed the entire seismicity of the IPOC network set up by G.A., S.B., J.-P.V. and B.S. S.H. performed the ETAS and  $b$ -value analysis. R.W., Y.Z. and T.D. contributed the co-seismic slip models. M.P. and F.T. performed the backprojection analysis. M.B. was responsible for the GPS data processing. J.B., A.H. and M.M. analysed geodetic locking and slip transients. A.H. modelled the accumulated, released and remaining moment. P.V. compiled the historical earthquake record, and O.O. wrote major parts of the mechanical interpretation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.S. ([schurr@gfz-potsdam.de](mailto:schurr@gfz-potsdam.de)).

# Dietary specializations and diversity in feeding ecology of the earliest stem mammals

Pamela G. Gill<sup>1</sup>, Mark A. Purnell<sup>2</sup>, Nick Crumpton<sup>1†</sup>, Kate Robson Brown<sup>3</sup>, Neil J. Gostling<sup>1†</sup>, M. Stampanoni<sup>4,5</sup> & Emily J. Rayfield<sup>6</sup>

The origin and radiation of mammals are key events in the history of life, with fossils placing the origin at 220 million years ago, in the Late Triassic period<sup>1</sup>. The earliest mammals, representing the first 50 million years of their evolution and including the most basal taxa, are widely considered to be generalized insectivores<sup>1,2</sup>. This implies that the first phase of the mammalian radiation—associated with the appearance in the fossil record of important innovations such as heterodont dentition, diphyodonty and the dentary–squamosal jaw joint<sup>1,3</sup>—was decoupled from ecomorphological diversification<sup>2,4</sup>. Finds of exceptionally complete specimens of later Mesozoic mammals have revealed greater ecomorphological diversity than previously suspected, including adaptations for swimming, burrowing, digging and even gliding<sup>2,5,6</sup>, but such well-preserved fossils of earlier mammals do not exist<sup>1</sup>, and robust analysis of their ecomorphological diversity has previously been lacking. Here we present the results of an integrated analysis, using synchrotron X-ray tomography and analyses of biomechanics, finite element models and tooth microwear textures. We find significant differences in function and dietary ecology between two of the earliest mammaliaform taxa, *Morganucodon* and *Kuehneotherium*—taxa that are central to the debate on mammalian evolution. *Morganucodon* possessed comparatively more forceful and robust jaws and consumed ‘harder’ prey, comparable to extant small-bodied mammals that eat considerable amounts of coleopterans. *Kuehneotherium* ingested a diet comparable to extant mixed feeders and specialists on ‘soft’ prey such as lepidopterans. Our results reveal previously hidden trophic specialization at the base of the mammalian radiation; hence even the earliest mammaliaforms were beginning to diversify—morphologically, functionally and ecologically. In contrast to the prevailing view<sup>2,4</sup>, this pattern suggests that lineage splitting during the earliest stages of mammalian evolution was associated with ecomorphological specialization and niche partitioning.

Recently, much progress has been made in understanding the pattern and timing of the radiation of mammals<sup>7–9</sup>, revealing successive waves of taxonomic and ecomorphological diversification in Middle–Late Jurassic to Palaeogene stem clades and crown groups<sup>2,10,11</sup>. However, understanding of early mammaliaforms and the initial radiation of mammals has lagged behind. Here we address this problem by testing the hypothesis that two of the earliest and most basal mammaliaforms were ecomorphologically distinct. *Morganucodon watsoni*<sup>12</sup> and *Kuehneotherium praecursoris*<sup>13</sup> are central to the debate on mammalian origins and are of fundamental phylogenetic importance (Extended Data Fig. 1). *Morganucodon* is one of the earliest (Late Triassic to Early Jurassic) and best-known Mesozoic mammals, with a global distribution; *Kuehneotherium* is of a similar age and size<sup>1,12,13</sup>. Both taxa are thought to be generalized insectivores<sup>1</sup> and co-existed (see Supplementary Information for discussion of sympatry) on a small landmass present during the Early Jurassic marine transgression (Hettangian–Early Sinemurian, about 200 Myr ago), in what is now Glamorgan, South Wales, UK<sup>1,12</sup> (Extended Data Fig. 2). In addition to the apomorphic mammalian jaw joint, both taxa retain the

plesiomorphic articular–quadrate jaw joint, as indicated by a well-developed postdentary trough (Fig. 1a, b), thus indicating that the postdentary bones still functioned as part of the jaw joint, rather than being incorporated into a definitive mammalian middle ear as in modern mammals<sup>1,2,12–14</sup> (*sensu ref. 15*). Curiously, *Kuehneotherium* possesses advanced molars, with cusps arranged in an obtuse-triangle pattern<sup>13,16</sup> (Extended Data Fig. 3b).

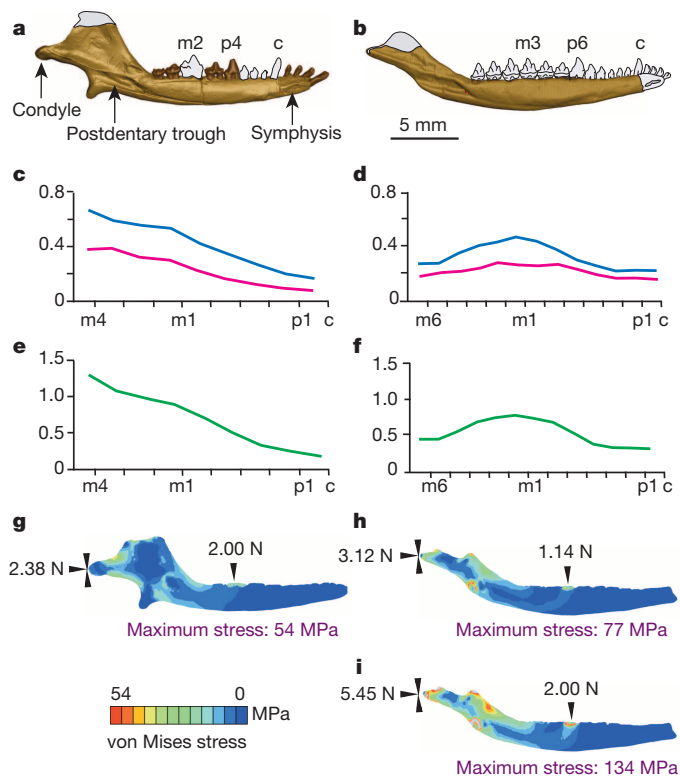
We tested hypotheses of functional and dietary specialization in these early mammaliaforms by generating digital mandibular reconstructions, and applying a suite of techniques: classical mechanics, finite element modelling and quantitative textural analysis of tooth microwear. The mandible is a good choice for study of feeding adaptations as it is primarily adapted for biting, and is not constrained by sensory systems such as eye or brain size<sup>17</sup>. Our null hypothesis was that functional performance did not differ between the two taxa.

Applying classical mechanics, we calculated the mechanical advantage for mid-molar, premolar and canine bites, reflecting the efficiency of the jaw system at transmitting force from the adductor muscles to the bite point. This revealed that *Morganucodon* has a notably larger mechanical advantage than *Kuehneotherium* (almost 50% greater during mid-molar biting) (Table 1), indicating that the mandible of *Morganucodon* had the potential to generate much larger bite forces than *Kuehneotherium*, and implying that *Kuehneotherium* bites were potentially faster but less forceful. We also determined jaw strength in bending and torsion during biting, treating the mandibular corpus as a beam<sup>18</sup>. The pattern of bending strength reveals a very different profile between the two taxa (Fig. 1c, d). *Morganucodon* shows peak resistance to bending at the rear of the tooth row as might be expected, as this region serves as a structural linkage between the tooth row and posterior functional elements of the jaw, such as the jaw joint and muscles<sup>19</sup>. However, *Kuehneotherium* shows peak resistance in the region of the anterior molars. Resistance to torsion (*J*) shows similar patterns (Fig. 1e, f). This different biomechanical profile in *Kuehneotherium* may reflect the importance of resisting bending in the central tooth row, to maintain the sharp bladed triangulated molars in precise occlusion<sup>16</sup>.

Finite element analysis allowed us to calculate stress, strain and deformation to assess the mechanical behaviour of the jaws<sup>20</sup>. This analysis can provide informative comparative data in the absence of known input parameters<sup>17</sup> and as such the two taxa were loaded with equal adductor muscle forces and constrained at the jaw joint and bite points (Extended Data Fig. 4). Finite element analysis shows that, during a simulated bite, despite similar length and surface area, the dentary of *Kuehneotherium* experiences greater maximum von Mises stress and maximum principal strain than *Morganucodon*, regardless of bite position, and higher reaction forces at the jaw joint, despite generating consistently less bite reaction force (Fig. 1g, h and Table 1). *Kuehneotherium* does not possess a robust condyle as in *Morganucodon* (Fig. 1a, b), further reducing its ability to withstand high reaction forces at the jaw joint. We tested whether *Morganucodon* or *Kuehneotherium* could generate enough bite force to

<sup>1</sup>School of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK. <sup>2</sup>Department of Geology, University of Leicester, University Road, Leicester LE1 7RH, UK.

<sup>3</sup>Department of Archaeology and Anthropology, University of Bristol, Woodland Road, Bristol BS8 1UU, UK. <sup>4</sup>Swiss Light Source, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. <sup>5</sup>Institute for Biomedical Engineering, University and ETH Zürich, Gloriastrasse 35, CH-8092 Zürich, Switzerland. <sup>6</sup>School of Earth Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK. †Present addresses: Natural History Museum, Cromwell Road, London SW7 5BD, UK (N.C.); Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK (N.J.G.).



**Figure 1 | Digital reconstructions and biomechanical analyses of *Morganucodon* and *Kuehneotherium* jaws.** **a, b,** Reconstructed left mandibles, medial view. Missing portions in grey. **a, M. watsoni.** **b, K. praecursoris;** dentition reconstructed from ref. 30. **c, d,** Section moduli (dorsoventral  $Z_x$  (blue) and mediolateral  $Z_y$  (red), in cubic millimetres) for (c) *Morganucodon* and (d) *Kuehneotherium*. **e, f,** Polar moment of inertia ( $J$ ) for (e) *Morganucodon* and (f) *Kuehneotherium*. **g–i,** Finite element von Mises stress contour plots for a unilateral molar bite (m2 for *Morganucodon* and m3 for *Kuehneotherium*) with closed jaw: **g,** *Morganucodon* and **h,** *Kuehneotherium*, with standard muscle loading; **i,** *Kuehneotherium* with muscle loading multiplied to give a bite reaction force of 2 N. Stress legend and contour plot scale apply to **g–i.** Single vertical arrows indicate tooth constraints; three grouped arrows indicate jaw joint constraints. (See Supplementary Information for links to scan data and finite element model images.)

pierce ‘hard’ insect cuticle (where ‘hard’ and ‘soft’ refer to the ease with which prey is pierced and chewed<sup>21</sup>). Estimation of bite force can circumscribe the range of potential prey, providing a measure of feeding performance and ecological partitioning<sup>22,23</sup>. A variety of insect prey was available at the time: the Glamorgan fissures have yielded beetle remains<sup>24</sup>, and soft-bodied insects, such as scorpion flies, were well established in the Early Jurassic<sup>25</sup>. (See Supplementary Information for discussion of potential prey.) A bite of 2 N is required to pierce the cuticle of a ‘hard’ insect (for example, a beetle) of appropriate prey size for *Morganucodon* or *Kuehneotherium*<sup>26,27</sup>. For *Morganucodon*, a simulated 2 N bite at mid-molar m2 (see Methods) did not generate excessive stress in the jaw (maximum 54 MPa) (Fig. 1g). For *Kuehneotherium*, increasing muscle loadings (keeping the ratio of muscle recruitment intact), to simulate a bite of 2 N at mid-molar m3 (Fig. 1i), produced higher reaction forces at the dentary condyle (5.45 N compared with 2.38 N for *Morganucodon*), and maximum von Mises stress values up to 134 MPa, which is 2.5 times that of *Morganucodon* and close to the value of tensile stress failure for bone<sup>28</sup>. This suggests that *Kuehneotherium* was probably incapable of processing ‘hard’ cuticle, and further illustrates differences in the biomechanical performance of the jaws. Comparative biomechanical data therefore point to morphofunctional and dietary specialization in these two taxa.

The hypothesis that *Morganucodon* and *Kuehneotherium* consumed different prey was independently tested by comparing their tooth microwear

**Table 1 | Biomechanical analysis results**

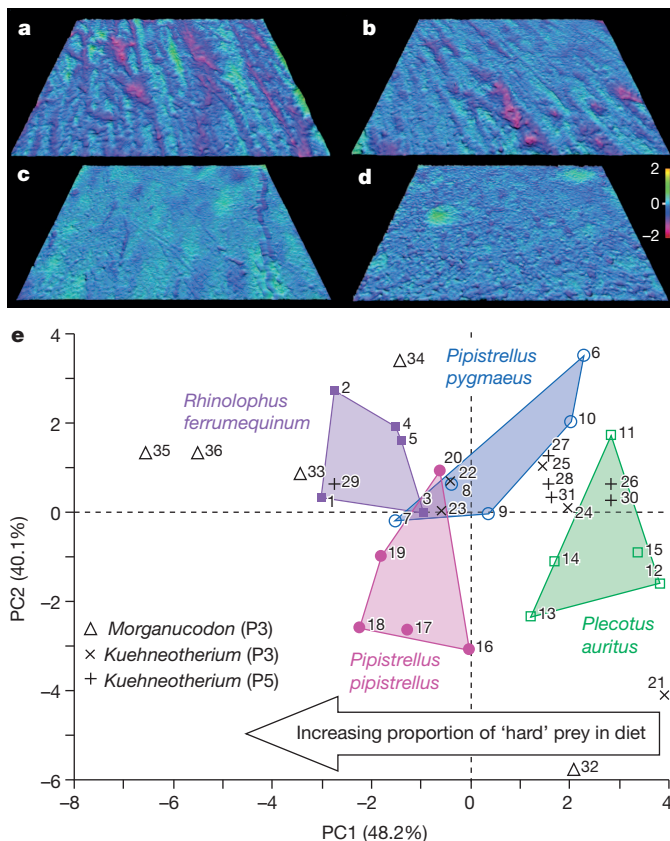
	<i>Morganucodon</i>	<i>Kuehneotherium</i>
<b>Mechanical advantage (three-dimensional)</b>		
Canine bite	0.51	0.31
Premolar bite	0.42	0.24
Molar bite	0.35	0.18
<b>Maximum von Mises stress (MPa)</b>		
Standard loading canine bite	66	97
Standard loading premolar bite	61	83
Standard loading molar bite	54	77
<b>Maximum principal strain (microstrain)</b>		
Standard loading canine bite	3,840	5,800
Standard loading premolar bite	3,540	5,020
Standard loading molar bite	3,100	4,440
<b>Reaction forces (N)</b>		
	<b>Jaw joint</b>	<b>Bite</b>
Standard loading canine bite	2.96	1.31
Standard loading premolar bite	2.69	1.62
Standard loading molar bite	2.38	2.00

Comparative mechanical advantage, maximum von Mises stress values (in megapascals), maximum principal strain values and reaction forces (in newtons) for *Morganucodon* and *Kuehneotherium* dentaries. See Methods section for an explanation of the standard loading for muscles. These loads achieve a 2 N reaction bite force at the molar (m2) of *Morganucodon*, sufficient to pierce insect cuticle.

textures with those of extant insectivores with known dietary preferences (specimens listed in Extended Data Table 1). Recent work on insectivorous bats has shown that microwear textural analysis based on three-dimensional roughness parameters discriminates between insectivore species that consume different proportions of ‘hard’ prey (such as beetles) and ‘soft’ prey (such as moths)<sup>29</sup>. Bats provide a useful comparative data set for our work because of their well-studied dietary differences and similarity in size to *Morganucodon* and *Kuehneotherium*. We compared the fossil taxa with four species of bats: *Plecotus auritus* (brown long-eared bat; a specialist on ‘soft’ insects); *Pipistrellus pipistrellus* (common pipistrelle) and *Pipistrellus pygmaeus* (soprano pipistrelle) (more mixed diet, both specialize on Diptera (flies), but *P. pipistrellus* consumes insects with a wider range of cuticle ‘hardness’ and more ‘hard’ prey than *P. pygmaeus*); and *Rhinolophus ferrumequinum* (greater horseshoe bat; mixed diet, but including more beetles—prey that is among the ‘hardest’ of insects) (see Extended Data Table 2 for dietary details). In the bats, nine roughness parameters differ significantly between species<sup>29</sup>, and principal component analysis (PCA) of these parameters (Fig. 2) separates bats according to dietary preferences in a space defined by principal component axes 1 and 2 (together accounting for 88.3% of variance), with axis 1 strongly correlated with dietary preferences ( $r_s = 0.81$ ,  $P < 0.0001$ ). Increasingly negative values indicate higher proportions of ‘hard’ prey, while increasingly positive values indicate increasing proportions of ‘soft’ prey<sup>29</sup>.

Projecting data for *Kuehneotherium* and *Morganucodon* onto the axes resulting from the analysis of bats produces clear separation of the two taxa. *Morganucodon* has negative values for principal component 1 (PC1), overlapping and extending beyond values for *R. ferrumequinum*. Slightly rougher textures in *Morganucodon* suggest that it consumed a higher proportion of ‘hard’ prey. Most *Kuehneotherium* specimens have positive values for PC1, overlapping the range of the ‘soft’ insect specialist *Pl. auritus*. Two specimens have negative PC1 values and plot into a space defined by the mixed-feeding *Pipistrellus*. Thirteen roughness parameters from *Morganucodon* and *Kuehneotherium* are correlated with the bat dietary axis (PC1; Extended Data Tables 3–5), including nine of the ten parameters that in bats are correlated with diet<sup>29</sup>, and values for PC1 differ significantly between the two fossil species ( $F = 5.67$ ; d.f. = 6, 29;  $P = 0.0005$ ). Pairwise tests (Tukey’s honestly significant difference;  $P < 0.05$ ) indicate that microwear textures in *Morganucodon* and *Kuehneotherium* differ from one another, yet *Morganucodon* does not differ from bats with mixed or ‘harder’ diets, and *Kuehneotherium* does not differ from the ‘soft’ insect specialist and mixed feeders. *Kuehneotherium* specimens from different fissure localities do not differ from one another (see Supplementary Information for specimen and fissure details). That *Kuehneotherium* and *Morganucodon* are so clearly separated by application of PCA based on extant bats with different diets provides powerful evidence that the two fossil taxa had diets that differed significantly in terms





**Figure 2 | Quantitative textural analysis of microwear in bats and fossil mammaliaforms.** a–d, Scale-limited roughness surfaces of *Morganucodon* (a; specimen 34), *R. ferrumequinum* (b; specimen 1), *Kuehneotherium* (c; specimen 24) and *Pl. auritus* (d; specimen 12); 146  $\mu\text{m} \times 110 \mu\text{m}$ ; contour vertical scale in micrometres. e, PCA of International Organization for Standardization (ISO) roughness parameters from bats and mammaliaforms. PCA based on data for bats only, with *Morganucodon* ( $n = 5$ ) and *Kuehneotherium* data (Pontalium 3 (P3)  $n = 5$ ; Pant 5 fissure (P5)  $n = 6$ ) projected onto the bat PCA axes. There are two anomalous specimens: a single *Kuehneotherium* specimen (29) has PC1 values similar to *R. ferrumequinum*, and one of the *Morganucodon* specimens (32) plots as an outlier to all other teeth analysed.

of prey 'hardness', and provides independent validation of distinctive mechanical behaviour and function revealed through our standard beam analysis and finite element modelling.

In summary, our analyses reveal previously hidden trophic diversity and niche partitioning at the base of the mammalian radiation, supporting a hypothesis of coupled lineage splitting and ecomorphological adaptation of the skull and jaws, even during the earliest stages of mammalian evolution. Our approach, combining biomechanical analyses with tooth microtextural validation of dietary differences, does not require exceptionally preserved specimens, and is applicable to fragmentary fossil remains. As such, it has the potential to provide direct evidence of ecomorphology and adaptation through a range of vertebrate radiations, using the most commonly preserved fossil elements: teeth and jaws.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 April; accepted 27 June 2014.

- Kielan-Jaworowska, Z., Cifelli, R. & Luo, Z.-X. *Mammals from the Age of Dinosaurs. Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
- Luo, Z.-X. Transformation and diversification in early mammal evolution. *Nature* **450**, 1011–1019 (2007).

- Kemp, T. S. *The Origin and Evolution of Mammals* Ch. 4 (Oxford Univ. Press, 2005).
- Bromham, L., Phillips, M. J. & Penny, D. Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends Ecol. Evol.* **14**, 113–118 (1999).
- Ji, Q., Luo, Z.-X., Yuan, C. X. & Tabrum, A. R. A swimming mammaliaform from the Middle Jurassic and ecomorphological diversification of early mammals. *Science* **311**, 1123–1127 (2006).
- Meng, J., Hu, Y., Wang, Y., Wang, X. & Li, C. A Mesozoic gliding mammal from northeastern China. *Nature* **444**, 889–893 (2006).
- Meredith, R. W. *et al.* Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
- dos Reis, M., Inoue, J., Asher, R. J., Donoghue, P. C. J. & Yang, Z. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B* **279**, 3491–3500 (2012).
- O'Leary, M. A. *et al.* The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–667 (2013).
- Wilson, G. P. *et al.* Adaptive radiation of multituberculate mammals before the extinction of dinosaurs. *Nature* **483**, 457–460 (2012).
- Grossnickle, D. M. & Polly, P. D. Mammal disparity decreases during the Cretaceous angiosperm radiation. *Proc. R. Soc. B* **280**, 20132110 (2013).
- Kermack, K. A., Mussett, F. & Rigney, H. W. Lower jaw of *Morganucodon*. *Zool. J. Linn. Soc.* **53**, 87–175 (1973).
- Kermack, D. M., Kermack, K. A. & Mussett, F. The Welsh pantothere *Kuehneotherium praecursoris*. *Zool. J. Linn. Soc.* **47**, 407–423 (1968).
- Luo, Z.-X. Developmental patterns in Mesozoic evolution of mammal ears. *Annu. Rev. Ecol. Evol. Syst.* **42**, 355–380 (2011).
- Allin, E. F. & Hopson, J. A. in *The Evolutionary Biology of Hearing* (eds Webster, D. B., Fay, R. R. & Popper, A. N.) 587–614 (Springer, 1992).
- Crompton, W. A. & Jenkins, F. A. American Jurassic symmetrodonts and Rhaetic pantotheres. *Science* **155**, 1006–1009 (1967).
- Tseng, Z. J., Mcnitt-Gray, J. L., Flashner, H., Wang, X. & Enciso, R. Model sensitivity and use of the comparative finite element method in mammalian jaw mechanics: mandible performance in the gray wolf. *PLoS ONE* **6**, e19171 (2011).
- Therrien, F. Mandibular force profiles of extant carnivorans and implications for the feeding behaviour of extinct predators. *J. Zool.* **267**, 249–270 (2005).
- Freeman, P. W. & Lemen, C. A. Simple predictors of bite force in bats: the good, the better and the better still. *J. Zool.* **282**, 284–290 (2010).
- Rayfield, E. J. Finite element analysis and understanding the biomechanics and evolution of living and fossil organisms. *Annu. Rev. Earth Planet. Sci.* **35**, 541–576 (2007).
- Evans, A. R. & Sanson, G. D. Biomechanical properties of insects in relation to insectivory: cuticle thickness as an indicator of insect 'hardness' and 'intractability'. *Aust. J. Zool.* **53**, 9–19 (2005).
- Dumont, E. R. & Herrel, A. The effects of gape angle and bite point on bite force in bats. *J. Exp. Biol.* **206**, 2117–2123 (2003).
- Aguirre, L. F., Herrel, A., Van Damme, R. & Matthyssens, E. Ecomorphological analysis of trophic niche partitioning in a tropical savannah bat community. *Proc. R. Soc. Lond. B* **269**, 1271–1278 (2002).
- Gardiner, B. G. New Rhaetic and Liassic beetles. *Palaeontology* **1**, 87–88 (1961).
- Grimaldi, D. & Engel, M. S. *Evolution of the Insects* 469–472 (Cambridge Univ. Press, 2005).
- Freeman, P. W. & Lemen, C. A. Using scissors to quantify hardness of insects: do bats select or size or hardness? *J. Zool.* **271**, 469–476 (2007).
- Aguirre, L. F., Herrel, A., van Damme, R. & Matthyssens, E. The implications of food hardness for diet in bats. *Funct. Ecol.* **17**, 201–212 (2003).
- Currey, J. D. *Bones: Structure and Mechanics* 58–60 (Princeton Univ. Press, 2002).
- Purnell, M. A., Crompton, N., Gill, P. G. & Rayfield, E. J. Within-guild dietary discrimination from 3-D textural analysis of tooth microwear in insectivorous mammals. *J. Zool.* **291**, 249–257 (2013).
- Gill, P. G. *Kuehneotherium from the Mesozoic Fissure Fillings of South Wales*. PhD thesis, Univ. Bristol (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. Armstrong, R. Asher, E. Bernard, P. Brewer, J. Bright, I. Corfe, A. Currant, A. Gill, T. Goddard, C. Hintermueller, J. Hooker, G. Jones, S. Lautenschlager, M. Lowe, F. Marone, F. Marx, C. Palmer, M. Pound, M. Ruecklin and J. Sibbick. This work was funded by Natural Environment Research Council grants NE/E010431/1 and NE/K01496X/1 to E.J.R. and P.G.G.; M.A.P. was supported by NE/G018189/1. Use of the Swiss Light Source, Paul Scherrer Institut, was supported by the European Commission 6th Framework Programme (RII3-CT-2004-506008).

**Author Contributions** E.J.R., P.G.G. and M.A.P. designed the study and wrote the paper; P.G.G., E.J.R., N.J.G. and M.S. collected the synchrotron radiation X-ray tomographic microscopy data; K.R.B. and P.G.G. collected the micro-computed tomography scan data; P.G.G. created the reconstructions and digital models, and analysed the biomechanical results; P.G.G. and E.J.R. interpreted the biomechanical results; P.G.G. prepared and acquired specimens for microwear analysis, M.A.P. and N.C. collected the microwear data, M.A.P. analysed and interpreted the microwear results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.G. (pam.gill@bristol.ac.uk) or E.R. (e.rayfield@bristol.ac.uk).

# The timing and spatiotemporal patterning of Neanderthal disappearance

Tom Higham<sup>1</sup>, Katerina Douka<sup>1</sup>, Rachel Wood<sup>1,2</sup>, Christopher Bronk Ramsey<sup>1</sup>, Fiona Brock<sup>1</sup>, Laura Basell<sup>3</sup>, Marta Camps<sup>4</sup>, Alvaro Arrizabalaga<sup>5</sup>, Javier Baena<sup>6</sup>, Cecillio Barroso-Ruiz<sup>7</sup>, Christopher Bergman<sup>8</sup>, Coralie Boitard<sup>9</sup>, Paolo Boscato<sup>10</sup>, Miguel Caparrós<sup>11</sup>, Nicholas J. Conard<sup>12,13</sup>, Christelle Draily<sup>14</sup>, Alain Froment<sup>15</sup>, Bertila Galván<sup>16</sup>, Paolo Gambassini<sup>10</sup>, Alejandro García-Moreno<sup>17,37</sup>, Stefano Grimaldi<sup>18</sup>, Paul Haesaerts<sup>19</sup>, Brigitte Holt<sup>20</sup>, Maria-Jose Iriarte-Chiapusso<sup>5</sup>, Arthur Jelinek<sup>21</sup>, Jesús F. Jordá Pardo<sup>22</sup>, José-Manuel Maíllo-Fernández<sup>22</sup>, Anat Marom<sup>1,23</sup>, Julià Maroto<sup>24</sup>, Mario Menéndez<sup>22</sup>, Laure Metz<sup>25</sup>, Eugène Morin<sup>26</sup>, Adriana Moroni<sup>10</sup>, Fabio Negrino<sup>27</sup>, Eleni Panagopoulou<sup>28</sup>, Marco Peresani<sup>29</sup>, Stéphane Pirson<sup>30</sup>, Marco de la Rasilla<sup>31</sup>, Julien Riel-Salvatore<sup>32</sup>, Annamaria Ronchitelli<sup>10</sup>, David Santamaría<sup>31</sup>, Patrick Semal<sup>33</sup>, Ludovic Slimak<sup>25</sup>, Joaquín Soler<sup>24</sup>, Narcís Soler<sup>24</sup>, Aritza Villaluenga<sup>17</sup>, Ron Pinhasi<sup>34</sup> & Roger Jacobi<sup>35,36,†</sup>

**The timing of Neanderthal disappearance and the extent to which they overlapped with the earliest incoming anatomically modern humans (AMHs) in Eurasia are key questions in palaeoanthropology<sup>1,2</sup>. Determining the spatiotemporal relationship between the two populations is crucial if we are to understand the processes, timing and reasons leading to the disappearance of Neanderthals and the likelihood of cultural and genetic exchange. Serious technical challenges, however, have hindered reliable dating of the period, as the radiocarbon method reaches its limit at ~50,000 years ago<sup>3</sup>. Here we apply improved accelerator mass spectrometry <sup>14</sup>C techniques to construct robust chronologies from 40 key Mousterian and Neanderthal archaeological sites, ranging from Russia to Spain. Bayesian age modelling was used to generate probability distribution functions to determine the latest appearance date. We show that the Mousterian ended by 41,030–39,260 calibrated years BP (at 95.4% probability) across Europe. We also demonstrate that succeeding ‘transitional’ archaeological industries, one of which has been linked with Neanderthals (Châtelperronian)<sup>4</sup>, end at a similar time. Our data indicate that the disappearance of Neanderthals occurred at different times in different regions. Comparing the data with results obtained from the earliest dated AMH sites in Europe, associated with the Uluzzian technocomplex<sup>5</sup>, allows us to quantify the temporal overlap between the two human groups. The results reveal a significant overlap of 2,600–5,400 years (at 95.4% probability). This has important implications for models seeking to explain the cultural, technological and biological elements involved in the replacement of Neanderthals by AMHs. A mosaic of populations in Europe during the Middle to Upper Palaeolithic transition suggests**

**that there was ample time for the transmission of cultural and symbolic behaviours, as well as possible genetic exchanges, between the two groups.**

European Palaeolithic sites contain the best evidence for the replacement of one human group (Neanderthals) by another (AMHs)<sup>1</sup>. The nature and process of the replacement, both in cultural and genetic terms, has been the focus of extensive research<sup>1,6,7</sup>. Recent studies of complete Neanderthal and modern human genomic sequences suggest that Neanderthals and AMHs interbred outside Africa<sup>7</sup>. This resulted in an introgression of 1.5–2.1% of Neanderthal-derived DNA<sup>8</sup>, or perhaps more<sup>9</sup>, in all modern non-African human populations. The analysis of three Neanderthal mitochondrial DNA (mtDNA) genomes from Denisova (Russian Altai), Vindija (Croatia) and Mezmaiskaya (Russian North Caucasus) indicates that the greatest amount of gene flow into non-African AMHs occurred after these Neanderthal populations had separated from each other<sup>8</sup>. At present it is not clear whether interbreeding occurred once or several times outside Africa<sup>10</sup>, or where it happened. After the interbreeding episode(s), Neanderthals and their distinctive material culture disappeared and were replaced across Eurasia by AMHs, but the precise timing of this has remained difficult to identify in the absence of a reliable chronological framework<sup>3</sup>.

Recent research has shown that radiocarbon ages have usually underestimated the true age of Palaeolithic remains, sometimes by several millennia<sup>3</sup>. This is due largely to problems in removing young carbon contamination from old organic samples at the limit of the <sup>14</sup>C method. The application of more rigorous chemical protocols<sup>11–13</sup> has recently resulted in improved reliability and accuracy. Several determinations

<sup>1</sup>Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology & the History of Art, University of Oxford, Oxford OX1 3QY, UK. <sup>2</sup>Research School for Earth Sciences, Australian National University, Canberra 0200, Australia. <sup>3</sup>School of Geography, Archaeology and Palaeoecology (GAP), Queen's University Belfast, Belfast BT7 1NN, UK. <sup>4</sup>School of Languages, Literatures and Cultures, College Park, 4102 Jiménez Hall, University of Maryland, Maryland 20742-4821, USA. <sup>5</sup>Research Team on Prehistory (IT-622-13), IKERBASQUE, University of the Basque Country (UPV-EHU), Tomás y Valiente Street, 01006 Vitoria-Gasteiz, Spain. <sup>6</sup>Departamento Prehistoria y Arqueología, Universidad Autónoma de Madrid, Campus Cantoblanco, 28049 Madrid, Spain. <sup>7</sup>Fundación Instituto de Investigación de Prehistoria y Evolución Humana, Plaza del Coso 1, 14900 Lucena, Córdoba, Spain. <sup>8</sup>URS, 525 Vine Street, Suite 1800, Cincinnati, Ohio 45202, USA. <sup>9</sup>8 rue des Sapins, 67100 Strasbourg, France.

<sup>10</sup>Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente, U.R. Preistoria e Antropologia, Università degli Studi di Siena, Via Laterina 8, 53100 Siena, Italy. <sup>11</sup>Département de Préhistoire, Muséum National d'Histoire Naturelle, 75013 Paris, France. <sup>12</sup>Abt. Ältere Urgeschichte und Quartärökologie, Universität Tübingen, Schloss Hohentübingen, 72070 Tübingen, Germany. <sup>13</sup>Tübingen Senckenberg Center for Human Evolution and Paleocology, Schloss Hohentübingen, 72070 Tübingen, Germany. <sup>14</sup>Service public de Wallonie, DGO4, Service de l'Archéologie, rue des Martyrs, 22, B-6700 Arlon, Belgium. <sup>15</sup>Laboratoire d'Éco-anthropologie et Ethnobiologie, Muséum de l'Homme, 17 place du Trocadéro, 75116 Paris, France. <sup>16</sup>Departamento de Prehistoria, Arqueología, Antropología e Historia Antigua, Universidad de La Laguna, Campus de Guajara, 38071 Tenerife, Spain. <sup>17</sup>Monrepos Archaeological Research Centre and Museum for Human Behavioural Evolution, Schloss Monrepos, D-56567 Neuwied, Germany. <sup>18</sup>Laboratorio di Preistoria 'B. Bagolini', Dipartimento di Lettere e Filosofia, Università degli Studi di Trento, via Tommaso Gar, 14 I-38122 Trento, Italy. <sup>19</sup>Institut Royal des Sciences Naturelles de Belgique, rue Vautier 29, B-1000 Brussels, Belgium. <sup>20</sup>Department of Anthropology, University of Massachusetts, 103 Machmer Hall, Amherst, Massachusetts 01003, USA. <sup>21</sup>School of Anthropology, Emil W. Haurly Building, University of Arizona, Tucson, Arizona 85721-0030, USA. <sup>22</sup>Departamento de Prehistoria y Arqueología, UNED. Paseo Senda del Rey 7, 20840, Madrid, Spain. <sup>23</sup>The Kimmel Center for Archaeological Science, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>24</sup>Àrea de Prehistòria, Universitat de Girona, pl. Ferrater Mora 1, 17071 Girona, Spain. <sup>25</sup>CNRS, UMR 5608, TRACES, Toulouse Jean Jaurès University, Maison de la Recherche, 5 Allées Antonio Machado, 31058 Toulouse, Cedex 9, France. <sup>26</sup>Department of Anthropology, Trent University, Life and Health Sciences Building Block C, 2140 East Bank Drive, Peterborough, Ontario K9J 7B8, Canada. <sup>27</sup>Dipartimento di Antichità, Filosofia e Storia, Università di Genova, Via Balbi 2, Genova I-16126, Italy. <sup>28</sup>Ephoreia of Paleolithic Archaeology of Northern Greece, Arditou 34B, Athens 11636, Greece. <sup>29</sup>Università di Ferrara, Dipartimento di Studi Umanistici, Sezione di Scienze Preistoriche e Antropologiche, Corso Ercole I d'Este 32, I-44100 Ferrara, Italy. <sup>30</sup>Service public de Wallonie, DGO4, Direction de l'Archéologie, rue des Brigades d'Irlande, 1, B-5100 Jambes, Belgium. <sup>31</sup>Departamento de Historia, Universidad de Oviedo, c/Teniente Alfonso Martínez, s/n, 33011 Oviedo, Spain. <sup>32</sup>Département d'Anthropologie, Université de Montréal, C. P. 6128, Succursale Centre-ville, Montréal, Québec H3T 1N8, Canada. <sup>33</sup>Service of Scientific Heritage, Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium. <sup>34</sup>UCD Earth Institute and School of Archaeology, University College Dublin, Belfield, Dublin 4, Ireland. <sup>35</sup>Department of Prehistory and Europe, Franks House, The British Museum, London N1 5QJ, UK. <sup>36</sup>The Natural History Museum, Cromwell Road, London SW7 5BD, UK. <sup>37</sup>The Cantabria International Institute for Prehistoric Research (IIIPC), University of Cantabria, Avda. Los Castros, s/n. 39005 Santander, Spain.

†Deceased.

that had previously supported late Neanderthal survival have been shown to be marked underestimates (for example, Vindija<sup>14</sup>, Zafarraya<sup>15</sup> and Mezmaiskaya<sup>16</sup>) and should be set to one side.

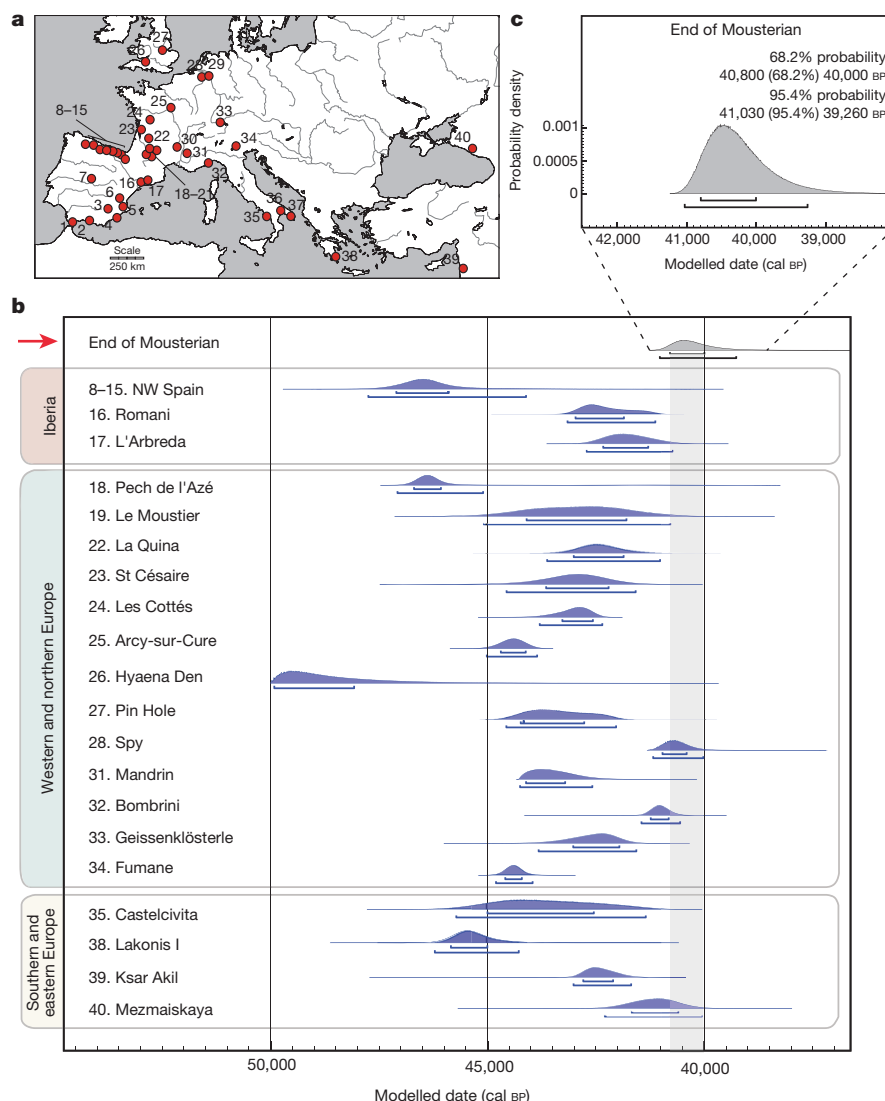
We performed extensive accelerator mass spectrometry (AMS) dating of critical late or final Mousterian archaeological horizons from 40 sites across Europe and the Mediterranean rim to explore the timing of Neanderthal extinction (Fig. 1a and Supplementary Methods). We also dated succeeding 'transitional' contexts, containing stone tool industries associated either with AMHs or with Neanderthals. These include Uluzzian (distributed across peninsular Italy and southern Greece and attributed to AMHs on the basis of associated AMH deciduous teeth excavated in Cavallo Cave<sup>5</sup>) and Châtelperronian (France and Cantabria Spain) layers, currently linked with Neanderthals on the basis of skeletal and technological evidence, although the association is debated<sup>17,18</sup>. Other transitional industries, such as the Szeletian and Bohunician of central and eastern Europe have not been dated as part of this study, nor have sites outside Europe, such as in the far northern Arctic fringes of Eurasia, where late Mousterian industries have been reported<sup>19</sup>.

We obtained 196 AMS radiocarbon measurements and used them to build high-precision age models using Bayesian statistics on the OxCal<sup>20</sup> platform. This allows us to incorporate stratigraphic and other relative age information, along with the calibrated likelihoods for each site. Probability distribution functions (PDFs) corresponding with the temporal boundaries of the latest Mousterian occupations were generated (Fig. 1b and Supplementary Methods).

The results show that the Mousterian end boundary PDFs all fall before 40,000 calibrated years (cal) BP (all probability ranges are expressed at 95.4%) (Fig. 1b). When placed into a single phase Bayesian model, the PDFs result in an overall end boundary ranging from 41,030–39,260 cal BP (Fig. 1c and Supplementary Methods). This PDF represents the age of the latest European Mousterian on the basis of our data.

The combined data suggest that the Mousterian ended at a very similar time, across sites ranging from the Black Sea and the Near East, to the Atlantic coast (Fig. 1a, b). Southern Iberia has been held to represent an exception to a wider European pattern<sup>21</sup>, with late survival of Neanderthals previously argued at sites such as Gorham's Cave, Gibraltar<sup>22</sup>. We could not reproduce any of the late dates from sites in this region<sup>15</sup> (Supplementary Methods) and it is apparent that many previous determinations underestimate the real age. It is unclear how long Neanderthals persisted in southern Iberia<sup>15</sup>. More dating evidence is required before we can determine whether Neanderthal presence was later here than elsewhere in Europe.

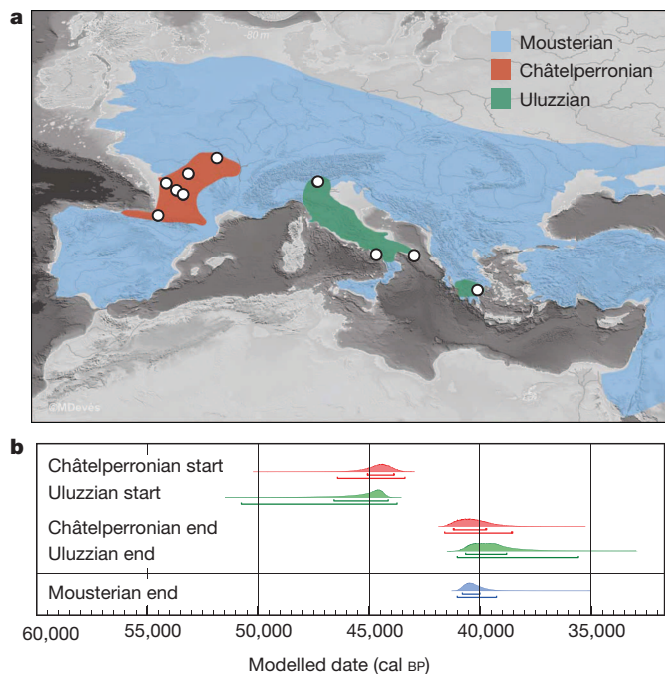
Our data also reveal differences in the spatiotemporal distribution of the latest Mousterian sites (Fig. 1b). The PDFs obtained were statistically ordered and the results show that significant differences exist between several late Mousterian contexts in different regions of Europe (Supplementary Methods). This may be attributed to the emergence of 'transitional' industries that replace the Mousterian between ~45,000–41,000 cal BP in some, but not all regions. At Fumane in Italy, for example, the Mousterian is replaced by the Uluzzian at 44,800–43,950 cal BP, while



**Figure 1 | Site locations and final boundary age ranges for Mousterian and Neanderthal sites**

**a**, Location of the 40 sites analysed and discussed in this paper. 1: Gorham's Cave; 2: Zafarraya; 3: El Niño; 4: Sima de las Palomas; 5: El Salt; 6: Quebrada; 7: Jarama VI; 8–15: La Viña, El Sidrón, La Güelga, Esquilleu, Morín, Arrillor, Labeko Koba, Lezetxiki; 16: Abric Romaní; 17: L'Arbreda; 18–21: Pech de l'Azé, Le Moustier, La Ferrassie, La Chappelle; 22: La Quina; 23: Saint-Césaire; 24: Les Cottés; 25: Arcy-sur-Cure; 26: Hyaena Den; 27: Pin Hole; 28: Spy; 29: Grotte Walou; 30: Néron; 31: Mandrin; 32: Bombrini/Mochi; 33: Geissenklösterle; 34: Fumane; 35: Castelcivita; 36: Oscuruscuto; 37: Cavallo; 38: Lakonis; 39: Ksar Akil; 40: Mezmaiskaya. **b**, Bayesian PDFs for the model boundaries of the final dated Mousterian phases by site across Europe (generated using OxCal4.2 software<sup>20</sup> and INTCAL13 (ref. 29)). **c**, PDF for the latest Mousterian based on the data in **b**.



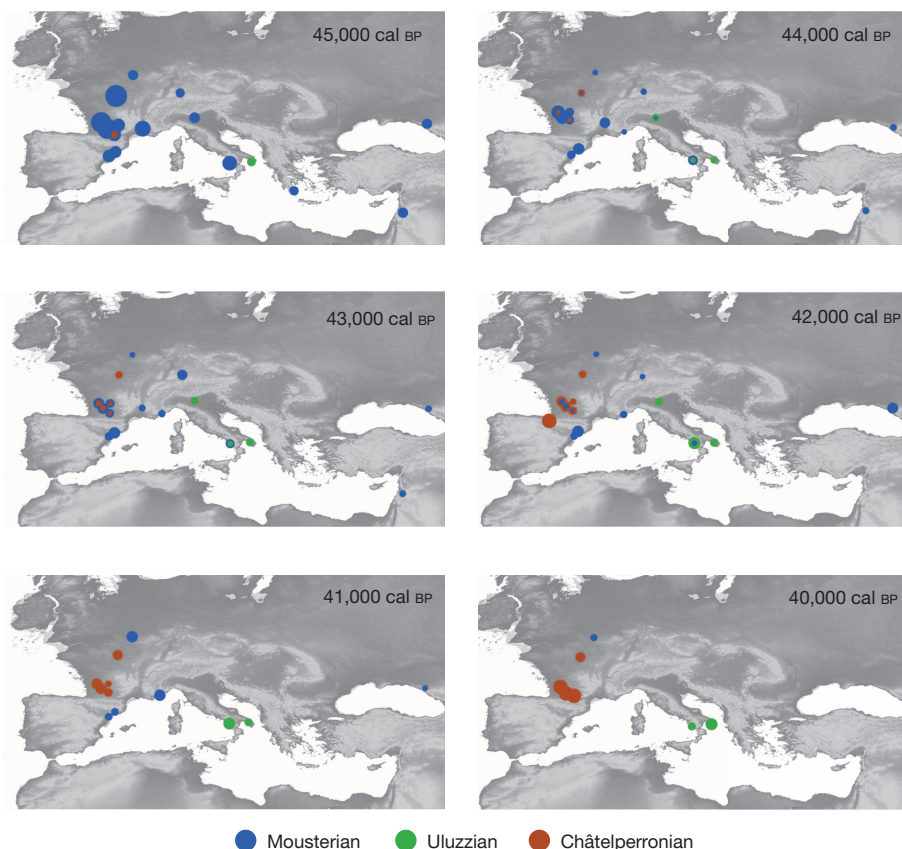


**Figure 2 | Transitional site locations and Bayesian age ranges for the start and end of the Châtelperronian and Uluzzian technocomplexes.** **a**, Geographic distribution of Châtelperronian (red), Uluzzian (green) and Mousterian (blue) technocomplexes. Map is shown with sea level at  $-80$  m below the present day<sup>1</sup>. Dated 'transitional' industry site locations are shown. Sea-level template map prepared by M. Devès. **b**, Bayesian modelled PDFs for the start and end boundaries of the Châtelperronian and Uluzzian in western Europe. The Mousterian end boundary (Fig. 1c) is shown for comparison. The three end boundaries overlap, but the late Mousterian always predates the two transitional industries stratigraphically where they co-occur.

at Mochi/Bombrini on the Italy–France border the Mousterian seems to last longer—until 41,460–40,500 cal BP. In the latter region, the Aurignacian arrives after a hiatus and no transitional complexes are evident. Since both the Uluzzian and Aurignacian are linked to AMHs, this lends support to the idea of a staggered replacement of Neanderthals in Italy as they neared local extinction (Supplementary Methods). Other late Mousterian contexts in sites in northern Spain, such as Abric Romani and L'Arbreda, are also considerably later than Fumane, suggesting that the Mousterian ended at different times in some parts of Europe.

The temporal range of the 'transitional' technocomplexes was also examined. With regard to the Châtelperronian, it is apparent on stratigraphic grounds that the Mousterian precedes it at all sites where both occur. However, our results show that the Châtelperronian at some sites (for example, Arcy-sur-Cure) starts statistically significantly before the end of the Mousterian at other sites in Europe such as Abric Romani and Geissenklösterle (Germany). If Neanderthals were responsible for both Mousterian and Châtelperronian, the implication is that there was considerable regional variation in their behaviour and adaptation strategies during this transition period. Assuming that the Châtelperronian is associated with Neanderthals, we combined the end boundaries for both into a single-phase Bayesian model and obtained a final 'Neanderthal' end PDF of 40,890–39,220 cal BP. The result is indistinguishable from the final Mousterian PDF, showing that uncertainty over the authorship of the Châtelperronian does not affect the age estimated for the last Neanderthals; they did not survive after  $\sim 41,000$ –39,000 cal BP (Fig. 2b).

By comparing the final Neanderthal PDF with those obtained for the start of the Uluzzian at the Cavallo site<sup>23</sup>, we can quantify the temporal overlap between Neanderthals and the earliest western European AMHs (Fig. 2b). The difference is significant and ranges from 2,600 to 5,400 years at 95.4% probability. Coexistence has been linked previously with the possibility of cultural transmission from AMHs to Neanderthals, termed 'acculturation'<sup>24</sup>, as a means of accounting for late Neanderthal technical and behavioural development. The early presence of AMHs in Mediterranean Europe by  $\sim 45,000$ –43,000 cal BP (ref. 23) and the



**Figure 3 | Time slices for western Europe between 45,000 and 40,000 cal BP showing the distribution of the Mousterian, Châtelperronian and Uluzzian modelled ages.** The size of the dots represents increasing and decreasing levels of the 95.4% probability ranges determined from the duration (date range) of each industry, as calculated by individual Bayesian site models (Supplementary Methods). Dots with two colours indicate overlapping date range probabilities for two industries found at the same site.

potential overlapping time may have acted as a stimulus for putative Neanderthal innovative and symbolic behaviour in the millennia before their disappearance. When we compare the start and end boundary PDFs for both Uluzzian and Châtelperronian sites we observe that they are very similar (Fig. 2b). This may provide further support for an acculturation model. Alternatively, this similarity in the start dates of the two industries might be seen as reflecting an AMH authorship for both. If this were the case, then the distribution of early modern humans would be wider than expected. Since the physical evidence linking these industries to different human groups is scarce, these interpretations are potentially prone to change with new excavation data.

High-precision chronometric data and Bayesian modelling allows us to map the spatiotemporal relationship between the three technocomplexes during the period ~45,000–41,000 cal BP as a series of time slices (Fig. 3 and Supplementary Methods). Since there is little to no robust evidence for interstratification of the transitional industries within Mousterian archaeological layers, we conclude that the chronological overlap observed must have also involved a degree of spatial separation between the two populations, regardless of whether Neanderthals were responsible for the Châtelperronian or not. In turn, this suggests that the dispersal of early AMHs was initially geographically circumscribed, proceeding step-wise, with the Uluzzian first and the Aurignacian following a few millennia later. The transitional industries, including those not analysed here, may be broadly contemporaneous technocomplexes that remained spatially distinct from one another. Rather than a rapid model of replacement of autochthonous European Neanderthals by incoming AMHs, our results support a more complex picture, one characterized by a biological and cultural mosaic that lasted for several thousand years.

## METHODS SUMMARY

AMS radiocarbon dating was undertaken at the Oxford Radiocarbon Accelerator Unit, University of Oxford. Collagen was extracted using the methods outlined previously<sup>11,25</sup>. Shell samples were dated according to the protocol outlined previously<sup>26</sup>. An acid–base oxidation/stepped combustion (ABOX-SC) method was used for charcoal<sup>13</sup>. Radiocarbon ages are given as conventional ages BP as described previously<sup>27</sup>. Corrections were made to bone collagen AMS determinations using a laboratory pre-treatment background subtraction<sup>28</sup>. Bones analysed range from very well preserved (a maximum of 14.9wt% collagen) to poorly preserved (a minimum of ~1.0wt% collagen). C:N atomic ratios and other analytical parameters were measured to determine the quality of the extracted collagen. The IntCal13 and Marine13 (ref. 29) calibration curves and the OxCal4.2 (ref. 20) program were used in the calibration and Bayesian age modelling. Supplementary Methods contains details of the archaeological sites investigated, the samples used, all determinations and the full Bayesian analysis.

Received 7 May; accepted 27 June 2014.

- Mellars, P. Neanderthals and the modern human colonization of Europe. *Nature* **432**, 461–465 (2004).
- d'Errico, F., Zilhão, J., Julien, M., Baffier, D. & Pelegrin, J. Neanderthal acculturation in western Europe? A critical review of the evidence and its interpretation. *Curr. Anthropol.* **39**, S1–S44 (1998).
- Higham, T. European Middle and Upper Palaeolithic radiocarbon dates are often older than they look: problems with previous dates and some remedies. *Antiquity* **85**, 235–249 (2011).
- Lévêque, F. & Vandermeersch, B. Découverte de restes humains dans le niveau Castelperronien à Saint-Césaire (Charente-Maritime). *C.R. Acad. Sc. Paris* **291**, 187–189 (1980).
- Benazzi, S. *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479**, 525–528 (2011).
- Villa, P. & Roebroeks, W. Neanderthal demise: an archaeological analysis of the modern human superiority complex. *PLoS ONE* **9**, e96424 (2014).
- Green, R. E. *et al.* A draft sequence of the Neanderthal genome. *Science* **328**, 710–722 (2010).

- Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
- Lohse, K. & Frantz, L. A. Neanderthal admixture in Eurasia confirmed by Maximum likelihood analysis of three genomes. *Genetics* **196**, 1241–1251 (2014).
- Wall, J. D. *et al.* Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199–209 (2013).
- Bronk Ramsey, C., Higham, T., Bowles, A. & Hedges, R. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* **46**, 155–163 (2004).
- Higham, T. G., Jacobi, R. M. & Bronk Ramsey, C. AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48**, 179–195 (2006).
- Bird, M. I. *et al.* Radiocarbon dating of 'old' charcoal using a wet oxidation-stepped combustion procedure. *Radiocarbon* **41**, 127–140 (1999).
- Higham, T., Ramsey, C. B., Karavanic, I., Smith, F. H. & Trinkaus, E. Revised direct radiocarbon dating of the Vindija G1 Upper Paleolithic Neanderthals. *Proc. Natl Acad. Sci. USA* **103**, 553–557 (2006).
- Wood, R. E. *et al.* Radiocarbon dating casts doubt on the late chronology of the Middle to Upper Palaeolithic transition in southern Iberia. *Proc. Natl Acad. Sci. USA* **110**, 2781–2786 (2013).
- Pinhasi, R. *et al.* Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc. Natl Acad. Sci. USA* **108**, 8611–8616 (2011).
- Higham, T. *et al.* The chronology of the Grotte du Renne (France) and implications for the association of ornaments and human remains within the Châtelperronian. *Proc. Natl Acad. Sci. USA* **107**, 20234–20239 (2010).
- Hublin, J.-J. *et al.* Radiocarbon dates from the Grotte du Renne and Saint-Césaire support a Neanderthal origin for the Châtelperronian. *Proc. Natl Acad. Sci. USA* **109**, 18743–18748 (2012).
- Slimak, L. *et al.* Late Mousterian persistence near the arctic circle. *Science* **332**, 841–845 (2011).
- Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
- Zilhão, J. & d'Errico, F. The chronology and taphonomy of the earliest Aurignacian and its implications for the understanding of Neanderthal extinction. *J. World Prehist.* **13**, 1–68 (1999).
- Finlayson, C. *et al.* Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* **443**, 850–853 (2006).
- Douka, K. *et al.* On the chronology of the Uluzzian. *J. Hum. Evol.* **68**, 1–13 (2014).
- Mellars, P. A. Major issues in the emergence of modern humans. *Curr. Anthropol.* **30**, 349–385 (1989).
- Brock, F. *et al.* Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon* **52**, 103–112 (2010).
- Douka, K. *et al.* Improved AMS <sup>14</sup>C dating of shell carbonates using high-precision X-Ray Diffraction (XRD) and a novel density separation protocol (CarDS). *Radiocarbon* **52**, 735–751 (2010).
- Stuiver, M. & Polach, H. Discussion: reporting of <sup>14</sup>C data. *Radiocarbon* **19**, 355–363 (1977).
- Wood, R. E., Bronk Ramsey, C. & Higham, T. F. G. Refining the ultrafiltration bone pretreatment background for radiocarbon dating at ORAU. *Radiocarbon* **52**, 600–611 (2010).
- Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** The Natural Environment Research Council (NERC) funded this work (NE/D014077/1). Additional funding was received from the Leverhulme Trust, through the Ancient Human Occupation of Britain (AHOB) project, the NRCF (NERC Radiocarbon Facility) programme, Keble College (Oxford) and the European Research Council. We thank our many collaborators and their excavation teams, and all staff at the Oxford Radiocarbon Accelerator Unit for their contribution to this work. Maps at ~80 m below current sea level were produced by M. Devès and A. Scheder Black.

**Author Contributions** T.H. and R.J. conceived the project. T.H. obtained funding and directed the project. T.H., R.W., K.D., F.B., C.B.R. and A.Ma. performed pre-treatment chemistry, AMS dating and Bayesian analysis using OxCal. T.H., R.W., K.D., L.B. and R.J. sampled materials for AMS dating. T.H. and K.D. wrote the paper and all co-authors contributed to the draft. K.D. and T.H. produced the figures and illustrations. R.J., L.B., M.C., A.A., J.B., C.B.-R., C.Be., C.Bo., P.B., M.C., N.J.C., C.D., A.F., B.G., P.G., A.G.-M., S.G., P.H., B.H., M.-J.-C., A.J., J.F.J.P., J.-M.M.-F., J.M., M.M., L.M., E.M., A.Mo., F.N., E.P., M.P., S.P., M.d.I.R., J.R.-S., A.R., D.S., P.S., L.S., J.S., N.S., A.V. and R.P. provided permits and archaeological samples, expertise on site sequences and prior data for the modelling.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H. ([thomas.higham@rlaha.ox.ac.uk](mailto:thomas.higham@rlaha.ox.ac.uk)).



# A microbial ecosystem beneath the West Antarctic ice sheet

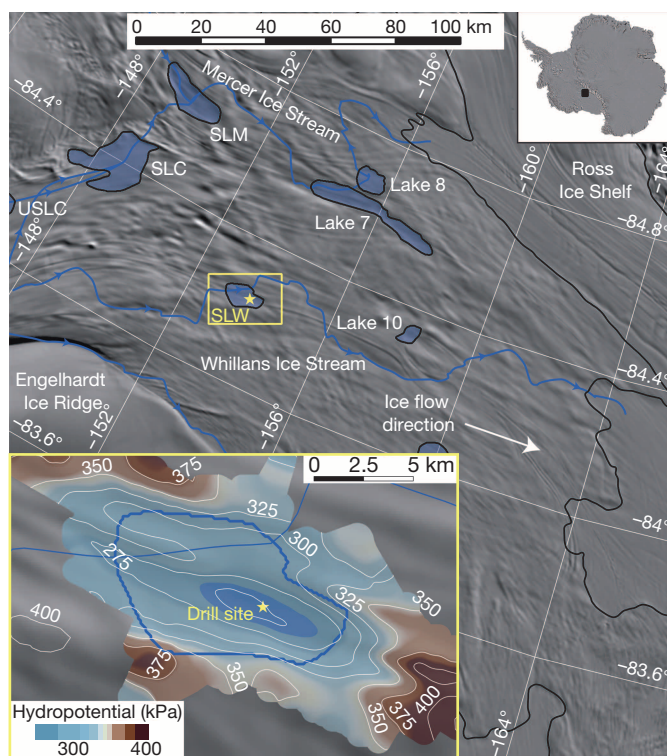
Brent C. Christner<sup>1</sup>, John C. Prisco<sup>2</sup>, Amanda M. Achberger<sup>1</sup>, Carlo Barbante<sup>3</sup>, Sasha P. Carter<sup>4</sup>, Knut Christianson<sup>5†</sup>, Alexander B. Michaud<sup>2</sup>, Jill A. Mikucki<sup>6</sup>, Andrew C. Mitchell<sup>7</sup>, Mark L. Skidmore<sup>8</sup>, Trista J. Vick-Majors<sup>2</sup> & the WISSARD Science Team<sup>‡</sup>

Liquid water has been known to occur beneath the Antarctic ice sheet for more than 40 years<sup>1</sup>, but only recently have these subglacial aqueous environments been recognized as microbial ecosystems that may influence biogeochemical transformations on a global scale<sup>2–4</sup>. Here we present the first geomicrobiological description of water and surficial sediments obtained from direct sampling of a subglacial Antarctic lake. Subglacial Lake Whillans (SLW) lies beneath approximately 800 m of ice on the lower portion of the Whillans Ice Stream (WIS) in West Antarctica and is part of an extensive and evolving subglacial drainage network<sup>5</sup>. The water column of SLW contained metabolically active microorganisms and was derived primarily from glacial ice melt with solute sources from lithogenic weathering and a minor seawater component. Heterotrophic and autotrophic production data together with small subunit ribosomal RNA gene sequencing and biogeochemical data indicate that SLW is a chemosynthetically driven ecosystem inhabited by a diverse assemblage of bacteria and archaea. Our results confirm that aquatic environments beneath the Antarctic ice sheet support viable microbial ecosystems, corroborating previous reports suggesting that they contain globally relevant pools of carbon and microbes<sup>2,4</sup> that can mobilize elements from the lithosphere<sup>6</sup> and influence Southern Ocean geochemical and biological systems<sup>7</sup>.

Almost 400 subglacial lakes have been identified beneath the Antarctic ice sheet<sup>8</sup>. Speculation on the presence of functional microbial ecosystems within these lakes followed their discovery<sup>1</sup> and motivated the initial studies of samples originating from Subglacial Lake Vostok (SLV)<sup>9,10</sup>. However, the body of microbiological data from SLV has been a point of contention, primarily because all studies were based on analyses of frozen (that is, accreted) lake water samples recovered from a borehole containing a contaminated hydrocarbon drilling fluid<sup>3</sup>. Our report documents the first analysis of water and surficial sediments collected directly from a subglacial lake beneath the West Antarctic ice sheet (WAIS) using microbiologically clean drilling and sampling techniques<sup>11</sup>.

The water residence time for SLV exceeds 10,000 years<sup>12</sup>, while that for 'active' lakes such as SLW is on the order of years to decades<sup>5,8</sup>. SLW is part of a network of three major reservoirs beneath the lower ice plain of the WIS that regulate water transport to a subglacial estuary at the grounding zone, linking the hydrological system to the sub-ice-ocean cavity beneath the Ross Ice Shelf<sup>5,13</sup> (Fig. 1). During two separate drainage events in 2006 and 2009, SLW discharged ~0.15 km<sup>3</sup> of water over two six-month periods, each time lowering the lake level by about 5 m<sup>5,14</sup>. The drilling location to access SLW was selected using reflection seismology<sup>13</sup> and ice-penetrating radar<sup>14</sup> data, and corresponded to the region of maximum predicted water column thickness, lowest hypopotential, and largest satellite-measured surface elevation changes (Fig. 1).

A hot water drilling system was used to create a ~0.6 m diameter bore-hole through the overlying ice sheet into SLW, allowing for physical measurements and the direct collection of water column and sediment samples. Drilling and lake entry procedures followed recommendations for environmental protection of subglacial aquatic environments<sup>11</sup>, incorporating rigorous measures to reduce the introduction of foreign microbiota and material into SLW and the interconnected subglacial drainage



**Figure 1 | Locator map of the WIS and SLW.** The yellow box and star indicate the general location of the lake and the drill site; maximum extent of SLW and other lakes<sup>28</sup> under the ice stream are shaded in blue; predicted subglacial water flowpaths through SLW and other subglacial lakes are represented by blue lines with arrows; the black line denotes the ice-sheet grounding line at the edge of the Ross Ice Shelf<sup>29</sup>. Inset (expanded from area in yellow box) shows details of SLW with both maximum (solid blue line) and minimum lake extent (shaded blue area), hypopotential contours (white isolines; 25 kPa interval), and drill site (yellow star; 84.240° S 153.694° W). Background imagery is MODIS MOA<sup>30</sup>.

<sup>1</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA. <sup>2</sup>Department of Land Resources and Environmental Science, Montana State University, Bozeman, Montana 59717, USA. <sup>3</sup>Institute for the Dynamics of Environmental Processes – CNR, Venice, and Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice 30123, Italy. <sup>4</sup>Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA. <sup>5</sup>Physics Department, St Olaf College, Northfield, Minnesota 55057, USA. <sup>6</sup>Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996, USA. <sup>7</sup>Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth SY23 3DB, UK. <sup>8</sup>Department of Earth Science, Montana State University, Bozeman, Montana 59717, USA. <sup>†</sup>Present address: Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA (K.C.).

<sup>‡</sup>Lists of participants and their affiliations appear at the end of the paper.



system. Video inspection of the borehole and temperature measurements revealed that the ice–water interface occurred at  $801 \pm 1$  m below the surface (mbs) and the lake depth at the borehole site was  $\sim 2.2$  m at the time of sampling. Two borehole deployments of a conductivity, temperature and depth (CTD) sonde together with data from three discrete hydrocasts showed that SLW had an average *in situ* temperature of  $-0.49$  °C, pH of 8.1, and conductivity of  $720 \mu\text{S cm}^{-1}$ ; properties that were distinctly different from the borehole water (Table 1).

Water from three discrete hydrocasts in SLW had near identical geochemical compositions on the basis of major ion chemistry (Table 1) and all showed oxygen under-saturation ( $\sim 16\%$  of air-saturated water). Since there is no definitive evidence of lake water freezing to the bottom of the overlying ice sheet as in SLV<sup>12</sup>, it is unlikely that lake water constituents in SLW are influenced significantly by freeze concentration. The  $\delta^{18}\text{O}$  of  $\text{H}_2\text{O}$  for SLW ( $-38.0\text{‰}$ ) was similar to glacial ice sampled approximately 10 m above the ice–water interface from the neighbouring Kamb Ice Stream<sup>15</sup> (KIS;  $-38$  to  $-39\text{‰}$ ), indicating that glacial melt

was the dominant water source for SLW. A considerable fraction of the major anions and cations originated from mineral weathering, with a minor seawater component based on  $\text{Cl}^-$  to  $\text{Br}^-$  ratios (Extended Data Table 1). Crustally derived non-seawater solutes in SLW showed a dominance of weathering products from silicate minerals ( $\text{Na}^+ + \text{K}^+$ ) over carbonate minerals ( $\text{Mg}^{2+} + \text{Ca}^{2+}$ ), similar to other sub ice-sheet systems in Greenland and Antarctica<sup>6,7</sup> (Supplementary Discussion). The dominant non-seawater anions ( $\text{SO}_4^{2-}$  and  $\text{HCO}_3^-$ ) were probably products of sulphide oxidation, carbonation reactions, and carbonate dissolution<sup>7</sup>. Sulphide oxidation and carbonation reactions have been demonstrated to be microbially driven in subglacial systems and linked to enhanced rates of mineral weathering<sup>16</sup>. Although clay minerals are a potential source of the relatively high  $\text{F}^-$  concentrations in SLW (Table 1), subglacial volcanism in the upstream catchment supplying SLW<sup>17</sup> may also contribute.

Ammonium accounted for 73% of the dissolved inorganic nitrogen pool within the water column of SLW (Table 1). Given that mineral sources of ammonium are minor, the majority of the ammonium is probably from microbial mineralization. Soluble reactive phosphorus levels were similar to the total inorganic nitrogen pool (dissolved N:P molar ratio of 1.1), implying a biologically nitrogen-deficient environment, relative to phosphorus. Unfortunately, sample limitations precluded measurement of dissolved organic N and P concentrations to assess their nutritional contribution. In addition to its nutritional role, ammonium is also an energy source for chemolithoautotrophic ammonium-oxidizing bacteria and archaea. Evidence for complete nitrification in the aerobic SLW water column was supported by  $\Delta^{17}\text{O}$  of  $\text{NO}_3^-$  values ( $-0.1\text{‰}$  to  $0.2\text{‰}$ ) that indicated microbial processes rather than atmospheric input was the dominant source for nitrate in the lake<sup>18</sup>. Particulate organic C (PC) to N (PN) molar ratios in the water column exceeded that of actively growing bacteria by almost 15-fold, suggestive of elevated levels of nitrogen-poor detritus. Dissolved organic carbon (DOC) in the water column averaged  $221 \pm 55 \mu\text{mol l}^{-1}$ , which is about five times greater than average values for the deep ocean<sup>19</sup> and similar to the maximum range estimate for SLV<sup>9,20</sup> ( $86$ – $160 \mu\text{mol l}^{-1}$ ). Acetate and formate concentrations in the water column averaged  $1.3$  and  $1.2 \mu\text{mol l}^{-1}$ , respectively, indicating that at least a portion of the DOC pool was labile. The conductivity and microbiological data (Table 1 and Fig. 2a) showed that little mixing occurred between the borehole water and lake, supporting the hypothesis that DOC in the water column originated from SLW. The lack of winnowing in sediment cores from SLW, in concert with the fact that similar DOC concentrations were obtained as the overlying ice moved  $\sim 4$  m during the course of our operations, provided evidence that water column DOC did not result from sediment disturbance during drilling operations. The DOC in SLW most likely originated from upward diffusion of DOC associated with ancient marine sediments<sup>4</sup> (SLW sediment surface area:depth ratio  $\approx 30,000$ ), chemoautotrophic production, or from a combination of both sources.

The average cell density in the SLW water column was  $1.3 \times 10^5$  cells  $\text{ml}^{-1}$  (Table 1); microscopy revealed the presence of numerous morphotypes, approximately 10% of which were filamentous (Fig. 3). Cellular ATP, a proxy for viable biomass, in SLW was  $3.7 \text{ pmol ATP l}^{-1}$  (Table 1). Cell and ATP concentrations were 188- and 93-fold higher, respectively, than those observed in the borehole water before breakthrough to SLW. Carbon biomass estimates for SLW water based on the ATP data ( $480 \pm 100 \text{ ng C l}^{-1}$ ) were 3- to 50-fold higher than those observed beneath the Ross Ice Shelf at site J9 (ref. 21). Analysis of small subunit ribosomal RNA (SSU rRNA) sequences amplified from the water column samples showed that the community was similar among replicate lake samples, was distinct from the drilling water (Fig. 2a), and contained at least 3,931 operational taxonomic units (OTUs; Extended Data Table 2). An OTU closely related to the nitrite oxidizing betaproteobacterium '*Candidatus Nitrotoga arctica*'<sup>22</sup> comprised 13% of the sequence data, and many of the most abundant phylotypes were closely related to chemolithoautotrophic species that use reduced nitrogen, iron or sulphur compounds as energy sources (Fig. 2b; Supplementary Discussion). Two of

**Table 1 | Biogeochemical data from the SLW borehole, water column, and surficial sediments**

Parameter	Borehole*	Water column†	Sediments‡
<b>Physical</b>			
Temperature (°C)§	$-0.17$ (0.25)	$-0.49$ (0.03)	n.d.
Conductivity ( $\mu\text{S cm}^{-1}$ @ 25 °C)	5.3	720 (10)	860
pH	5.4	8.1 (0.1)	7.3
Redox (mV (SHE))	n.d.	382	n.d.
<b>Microbiological</b>			
Cell density (cells $\text{ml}^{-1}$ )	$6.9 \times 10^2$ (51.0)	$1.3 \times 10^5$ ( $0.4 \times 10^5$ )	n.d.
Cellular ATP ( $\text{pmol l}^{-1}$ )	0.04 (0.002)	3.70 (1.00)	n.d.
[ <sup>3</sup> H]thymidine¶	n.d.	13.7 (1.3)	46.6 (5.6)
[ <sup>3</sup> H]leucine¶	n.d.	2.9 (0.4)	0.9 (0.04)
<sup>14</sup> C-bicarbonate ( $\text{ng C l}^{-1} \text{ d}^{-1}$ )	n.d.	32.9 (4.2)	n.d.
<b>Carbon and nutrients</b>			
Dissolved oxygen ( $\mu\text{mol l}^{-1}$ )	n.d.	71.9 (12.5)	n.d.
DIC ( $\text{mmol l}^{-1}$ )	n.d.	2.11 (0.03)	n.d.
DOC ( $\mu\text{mol l}^{-1}$ )	n.d.	221 (55)	n.d.
Acetate ( $\mu\text{mol l}^{-1}$ )	n.d.	1.3 (0.2)	n.d.
Formate ( $\mu\text{mol l}^{-1}$ )	n.d.	1.2 (0.3)	n.d.
PC#	n.d.	78.5 (7.4)	384.2 (37.0)
PN#	n.d.	1.2 (0.4)	21.5 (1.7)
PC:PN (molar)	n.d.	65.4 (0.3)	17.9 (0.4)
$\text{NH}_4^+$ ( $\mu\text{mol l}^{-1}$ )	n.d.	2.4 (0.6)	n.d.
$\text{NO}_2^-$ ( $\mu\text{mol l}^{-1}$ )	n.d.	0.1 (0.1)	n.d.
$\text{NO}_3^-$ ( $\mu\text{mol l}^{-1}$ )	n.d.	0.8 (0.5)	9.1
$\text{PO}_4^{3-}$ ( $\mu\text{mol l}^{-1}$ )	n.d.	3.1 (0.7)	7.3
DIN:SRP (molar)	n.d.	1.1 (0.4)	n.d.
<b>Major ions (<math>\mu\text{eq l}^{-1}</math>)</b>			
$\text{Na}^+$	n.d.	5,276 (18)	6,977
$\text{K}^+$	n.d.	186 (4.2)	293 (1.0)★
$\text{Mg}^{2+}$	n.d.	507 (12)	596 (101)★
$\text{Ca}^{2+}$	n.d.	859 (29)	860 (104)★
$\text{F}^-$	n.d.	31.5 (0.4)	34.0
$\text{Cl}^-$	n.d.	3,537 (3.4)	4,943
$\text{Br}^-$	n.d.	6 (0.01)	7 (0.4)★
$\text{SO}_4^{2-}$	n.d.	1,111 (0.4)	1,230
$\text{HCO}_3^-$	n.d.	2,111 (35)	2,238**
<b>Stable isotopes††</b>			
$\delta^{18}\text{O}$ of $\text{H}_2\text{O}$	n.d.	$-38.0\text{‰}$	$-37.5\text{‰}$
$\Delta^{17}\text{O}$ of $\text{NO}_3^-$	n.d.	$-0.1$ to $0.2\text{‰}$	n.d.

\* Borehole water sampled by hydrocast at 672 mbs before lake entry.

† Water column data represent averages ( $\pm$  s.d.) from hydrocasts collected on 28 January 2013 (cast 1), 30 January (cast 2) and 31 January (cast 3) 2013, except for [<sup>3</sup>H]leucine incorporation, which is an average of cast 1 and 3 only.

‡ The sediment data correspond to measurements from the upper 2 cm of surficial sediments.

§ Average ( $\pm$  s.d.) of *in situ* measurements made through the lake water column at  $\sim 10$  cm intervals with a SBE 19plusV2 SeaCAT Profiler CTD on 28 January and 30 January 2013.

|| Based on measurements from discrete water samples brought to the surface.

¶ Macromolecular incorporation rates of tritium were converted to cellular carbon and presented along with bicarbonate incorporation as average  $\text{ng C l}^{-1} \text{ d}^{-1}$  ( $\pm$  s.d.) for water or average  $\text{ng C d}^{-1}$  gram dry weight<sup>-1</sup> ( $\pm$  s.d.) of sediment.

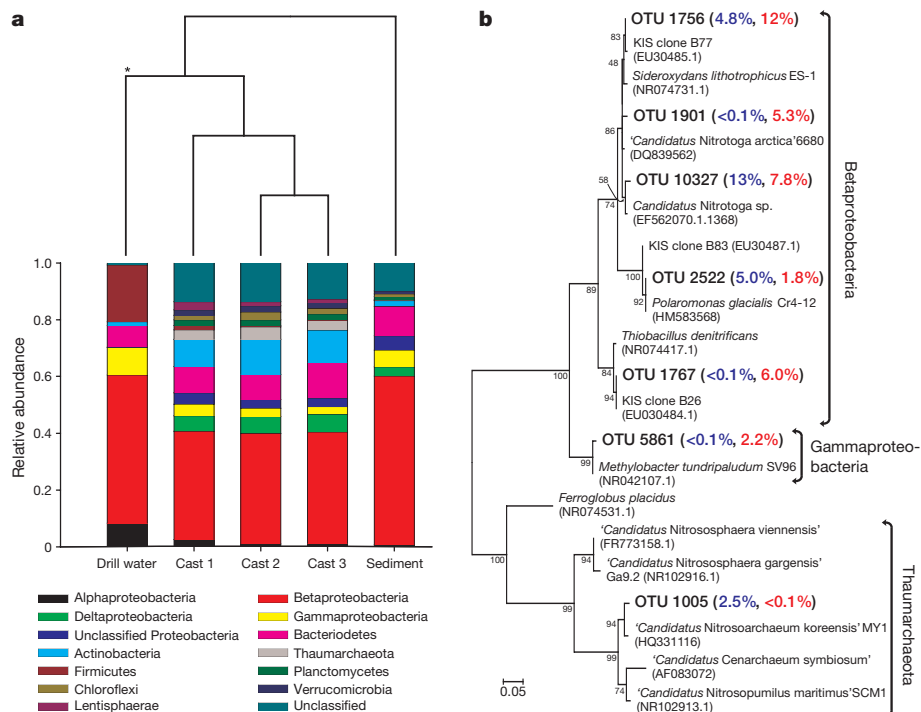
# Average ( $\pm$  s.d.)  $\mu\text{mol l}^{-1}$  for water and average ( $\pm$  s.d.)  $\mu\text{mol g dry weight}^{-1}$  for surficial sediment.

★ Surficial sediment porewater major ions are the average ( $\pm$  range) of two replicates.

\*\* Calculated based on charge balance.

†† Values are per thousand and reported relative to V-SMOW. The range of 2 measurements is given for  $\Delta^{17}\text{O}$  of  $\text{NO}_3^-$ .

n.d., no data available.



**Figure 2 | Phylogenetic analysis of SSU gene sequences obtained from the SLW water column, surficial sediment (0–2 cm) and drilling water.** **a**, Cluster analysis of the microbial phylogenetic structure in the samples (top) and the relative abundance of bacterial and archaeal phyla in the water and sediment samples (bottom). The Proteobacteria were split into classes for greater detail. The asterisk indicates statistical significance (analysis of molecular variance, AMOVA,  $P$  value < 0.001). **b**, Phylogenetic analysis of

bacterial and archaeal OTUs abundant in the SLW water column and sediments. The accession numbers of nearest neighbours and reference taxa are listed parenthetically. Bootstrap values are shown at the nodes. SLW phylotypes are bolded and followed by the percentage each represented in the water column (blue) and sediment (red) libraries. The scale bar indicates the number of nucleotide substitutions per position.

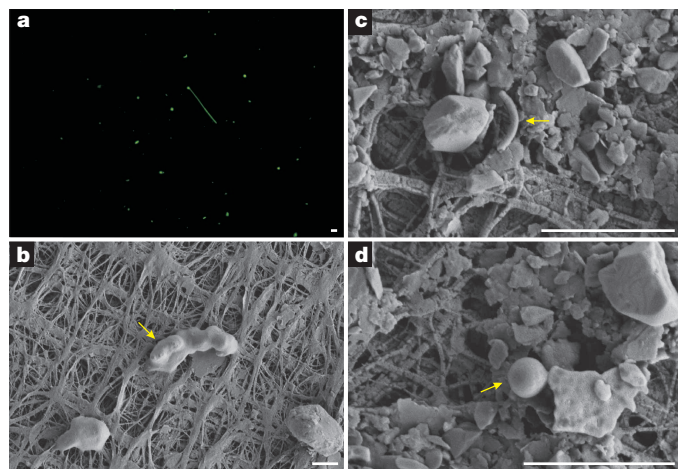
the abundant water column OTUs had high identity (>99%) to SSU sequences previously reported from sediments sampled beneath the KIS<sup>23</sup> (Fig. 2b). Preliminary attempts to detect eukaryotic SSU sequences in the SLW water column were unsuccessful.

Average dark [<sup>14</sup>C]bicarbonate incorporation in the water column samples (32.9 ng C l<sup>-1</sup> d<sup>-1</sup>; Table 1) exceeded average rates of heterotrophic production based on [<sup>3</sup>H]thymidine (13.7 ng C l<sup>-1</sup> d<sup>-1</sup>) and [<sup>3</sup>H]leucine (2.9 ng C l<sup>-1</sup> d<sup>-1</sup>) incorporation by 2- and 11-fold, respectively. Assuming that the thymidine and leucine values represent net incorporation, and that respiratory losses were 87% of net incorporation (which

are average values for Antarctic McMurdo Dry Valley lakes<sup>24</sup>), the gross bacterial carbon demand (net productivity + respiration) would be 105 and 23 ng C l<sup>-1</sup> d<sup>-1</sup>, respectively. If dark [<sup>14</sup>C]bicarbonate incorporation represents new organic carbon production via chemoautotrophy, the observed rates would meet between 31% and 143% of the heterotrophic carbon demand in the system. It should be noted that the effect of pressure (~8 MPa in SLW) was not tested and may influence the absolute rates of metabolism measured.

Pore water conductivity (860 μS cm<sup>-1</sup>) and pH (7.3) in SLW's surficial sediments were within 20% of the lake water values (Table 1). Upward diffusion of ions from sediment pore water is presumably the primary source of the ions in the water column. Average surficial sediment PC and PN concentrations were 384.2 and 21.5 μmol g dry weight<sup>-1</sup>, respectively, and represented 0.43% and 0.03% of sediment dry weight. The molar PC:PN ratio in the surficial sediment layer (17.9) was 3.7-fold lower than that in the water column (Table 1), indicative of nitrogen-enriched sedimentary particulate organic matter, with respect to water column suspensoids. On the basis of rates of thymidine and leucine incorporation, average heterotrophic production in the surficial sediment was 46.6 and 0.9 ng C d<sup>-1</sup> g dry weight<sup>-1</sup>, respectively. Approximately 75% of the OTUs from the surficial sediments classified within the Proteobacteria (Fig. 2a). Although many phylotypes in the water column were also abundant in the surficial sediments (Fig. 2b), ~70% of the OTUs were unique to the sediment environment. The nearest neighbours of the most abundant phylotypes in the surface sediments were chemolithoautotrophs or species that use C1 hydrocarbons as carbon and energy sources (Fig. 2b, Supplementary Discussion).

Our data show that SLW supports a metabolically active and phylogenetically diverse ecosystem that functions in the dark at sub-zero temperatures, confirming more than a decade of circumstantial evidence regarding the presence of life beneath Antarctica's ice sheet<sup>9,10,20,23</sup>. Rate experiments revealed that chemoautotrophic primary production in SLW



**Figure 3 | Morphological diversity of microbial cells in the SLW water column.** **a**, Epifluorescence micrograph showing a variety of cell morphotypes, which was confirmed by scanning electron microscopy (SEM; **b–d**). The yellow arrows in the SEM images indicate cells with rod (**b**), curved rod (**c**) and coccoid (**d**) morphologies. Scale bar, 2 μm.

is adequate to support heterotrophic metabolism in the subglacial ecosystem. The abundance of taxa related to nitrifiers<sup>22,25</sup> in concert with elevated ammonium and  $\Delta^{17}\text{O}$  of  $\text{NO}_3$  values near 0‰ in the water column (Table 1) implies that nitrification may be a fundamental chemotrophic pathway of new organic carbon production in SLW. Similar conclusions regarding the ecological significance of nitrification have been drawn for the water column beneath the Ross Ice Shelf<sup>26</sup> and in McMurdo Sound<sup>27</sup>. Given the prevalence of subglacial water in Antarctica<sup>8</sup>, our data from SLW lead us to contend that aquatic microbial ecosystems are common features of the subsurface environment that exists beneath the  $\sim 10^7 \text{ km}^2$  Antarctic ice sheet.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper

**Received 2 April; accepted 9 July 2014.**

- Oswald, G. K. A. & De Robin, G. Q. Lakes beneath the Antarctic ice sheet. *Nature* **245**, 251–254 (1973).
- Priscu, J. C. et al. in *Polar Lakes and Rivers* (eds Vincent, W. & Laybourn-Parry, J.) Ch.7 (Oxford Univ. Press, 2008).
- Christner, B. C., Skidmore, M. L., Priscu, J. C., Tranter, M. & Foreman, C. M. in (eds Margesin, R., Schinner, F., Marx, J.-C. & Gerday, C.) *Psychrophiles: From Biodiversity to Biotechnology* pp. 51–71 (Springer, 2008).
- Wadham, J. L. et al. Potential methane reservoirs beneath Antarctica. *Nature* **488**, 633–637 (2012).
- Fricker, H. A., Scambos, T., Bindaschadler, R. & Padman, L. An active subglacial water system in West Antarctica mapped from space. *Science* **315**, 1544–1548 (2007).
- Skidmore, M., Tranter, M., Tulaczyk, S. & Lanoil, B. Hydrochemistry of ice stream beds—evaporitic or microbial effects? *Hydrol. Processes* **24**, 517–523 (2010).
- Wadham, J. L. et al. Biogeochemical weathering under ice: size matters. *Glob. Biogeochem. Cycles* **24**, GB3025 (2010).
- Wright, A. & Siegert, M. A fourth inventory of Antarctic subglacial lakes. *Antarct. Sci.* **24**, 659–664 (2012).
- Priscu, J. C. et al. Geomicrobiology of subglacial ice above Lake Vostok. *Science* **286**, 2141–2144 (1999).
- Karl, D. M. et al. Microorganisms in the accreted ice of Lake Vostok. *Science* **286**, 2144–2147 (1999).
- Priscu, J. C. et al. A microbiologically clean strategy for access to the Whillans Ice Stream subglacial environment. *Antarct. Sci.* **25**, 637–647 (2013).
- Bell, R. E. et al. Origin and fate of Lake Vostok water frozen to the base of the East Antarctic ice sheet. *Nature* **416**, 307–310 (2002).
- Horgan, H. J. et al. Estuaries beneath ice sheets. *Geology* **41**, 1159–1162 (2013).
- Christianson, K., Jacobel, R. W., Horgan, H. J., Anandakrishnan, S. & Alley, R. B. Subglacial Lake Whillans—ice-penetrating radar and GPS observations of a shallow active reservoir beneath a West Antarctic ice stream. *Earth Planet. Sci. Lett.* **331–332**, 237–245 (2012).
- Vogel, S. W. et al. Subglacial conditions during and after stoppage of an Antarctic ice stream: is reactivation imminent? *Geophys. Res. Lett.* **32**, L14502 (2005).
- Montross, S. N., Skidmore, M., Tranter, M., Kivimäki, A.-L. & Parkes, R. J. A microbial driver of chemical weathering in glaciated systems. *Geology* **41**, 215–218 (2013).
- Blankenship, D. D. et al. Active volcanism beneath the West Antarctic ice-sheet and implications for ice-sheet stability. *Nature* **361**, 526–529 (1993).
- Michalski, G., Bhattacharya, S. K. & Girsch, G.  $\text{NO}_x$  cycle and tropospheric ozone isotope anomaly: an experimental investigation. *Atmos. Chem. Phys. Discuss.* **13**, 9443–9483 (2013).
- Hansell, D. A. & Carlson, C. A. Deep-ocean gradients in the concentration of dissolved organic carbon. *Nature* **395**, 263–266 (1998).
- Christner, B. C. et al. Limnological conditions in Subglacial Lake Vostok, Antarctica. *Limnol. Oceanogr.* **51**, 2485–2501 (2006).
- Azam, F. et al. Occurrence and metabolic activity of organisms under the Ross Ice Shelf, Antarctica, at Station J9. *Science* **203**, 451–453 (1979).
- Alawi, M., Lipski, A., Sander, T., Pfeiffer, E.-M. & Spieck, E. Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. *ISME J.* **1**, 256–264 (2007).
- Lanoil, B. et al. Bacteria beneath the West Antarctic ice sheet. *Environ. Microbiol.* **11**, 609–615 (2009).
- Takacs, C., Priscu, J. & McKnight, D. Bacterial dissolved organic carbon demand in McMurdo Dry Valley Lakes, Antarctica. *Limnol. Oceanogr.* **46**, 1189–1194 (2001).
- Walker, C. B. et al. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl Acad. Sci. USA* **107**, 8818–8823 (2010).
- Horrigan, S. G. Primary production under the Ross Ice Shelf, Antarctica. *Limnol. Oceanogr.* **26**, 378–382 (1981).
- Priscu, J. C., Downes, M. T., Priscu, L. R., Palmisano, A. C. & Sullivan, C. W. Dynamics of ammonium oxidizer activity and nitrous oxide ( $\text{N}_2\text{O}$ ) within and beneath Antarctic sea ice. *Mar. Ecol. Prog. Ser.* **62**, 37–46 (1990).
- Fricker, H. A. & Scambos, T. Connected subglacial lake drainage activity on lower Mercer and Whillans Ice Streams, West Antarctica, 2003–2008. *J. Glaciol.* **55**, 303–315 (2009).
- Depoorter, M. A. et al. Calving fluxes and basal melt rates of Antarctic ice shelves. *Nature* **502**, 89–92 (2013).
- Haran, T., Bohlander, J., Scambos, T. & Fahnestock, M. MODIS mosaic of Antarctica (MOA) image map. <http://dx.doi.org/10.7265/N5ZK5DM5> (National Snow and Ice Data Center, 2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The Whillans Ice Stream Subglacial Access Research Drilling (WISSARD) project was funded by National Science Foundation grants (0838933, 0838896, 0838941, 0839142, 0839059, 0838885, 0838855, 0838763, 0839107, 0838947, 0838854, 0838764 and 1142123) from the Division of Polar Programs. Partial support was also provided by funds from NSF award 1023233 (B.C.C.), NSF award 1115245 (J.C.P.), the NSF's Graduate Research Fellowship Program (1247192; A.M.A.), the Italian National Antarctic Program (C.B.), and fellowships from the NSF's IGERT Program (0654336) and the Montana Space Grant Consortium (A.B.M.). Logistics were provided by the 139th Expeditionary Airlift Squadron of the New York Air National Guard, Kenn Borek Air, and by many dedicated individuals working as part of the Antarctic Support Contractor, managed by Lockheed-Martin. The drilling was directed by F. Rack; D. Blythe, J. Burnett, C. Carpenter, D. Duling (chief driller), D. Gibson, J. Lemery, A. Melby and G. Roberts provided drill support at SLW. L. Geng, B. Vandenheuvel, A. Schauer and E. Steig provided assistance with the stable isotopic analyses. We thank J. Dore for assistance with the nutrient analysis.

**Author Contributions** The manuscript was written by B.C.C. and J.C.P.; A.M.A. generated and analysed the molecular data; C.B., A.C.M. and M.L.S. conducted and interpreted the chemical measurements; S.P.C. and K.C. provided geophysical data; J.A.M. obtained and examined the CTD data; A.B.M. and T.J.V. contributed and analysed physiological and biogeochemical data; M.L.S. conducted and interpreted the isotopic analyses; and T.J.V. provided the micrographs. All authors contributed to the study design and acquisition of samples and/or data.

**Author Information** The SSU sequence data are deposited in the NCBI SRA database under the accession number SRP041285. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.C.C. (xner@lsu.edu) or J.C.P. (jpriscu@montana.edu).

#### WISSARD Science Team Members

W. P. Adkins<sup>1</sup>, S. Anandakrishnan<sup>2</sup>, G. Barcheck<sup>3</sup>, L. Beem<sup>3</sup>, A. Behar<sup>4</sup>, M. Beitch<sup>3</sup>, R. Bolsey<sup>3</sup>, C. Branecky<sup>3</sup>, R. Edwards<sup>5</sup>, A. Fisher<sup>3</sup>, H. A. Fricker<sup>6</sup>, N. Foley<sup>3</sup>, B. Guthrie<sup>7</sup>, T. Hodson<sup>7</sup>, R. Jacobel<sup>8</sup>, S. Kelley<sup>5</sup>, K. D. Mankoff<sup>3</sup>, E. McBryan<sup>4</sup>, R. Powell<sup>7</sup>, A. Purcell<sup>9</sup>, D. Sampson<sup>3</sup>, R. Scherer<sup>7</sup>, J. Sherve<sup>5</sup>, M. Siegfried<sup>6</sup> & S. Tulaczyk<sup>3</sup>

<sup>1</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA. <sup>2</sup>Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>3</sup>Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, California 95064, USA. <sup>4</sup>School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA. <sup>5</sup>Department of Land Resources and Environmental Science, Montana State University, Bozeman, Montana 59717, USA. <sup>6</sup>Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA. <sup>7</sup>Department of Geology and Environmental Geosciences, Northern Illinois University, DeKalb, Illinois 60115, USA. <sup>8</sup>Physics Department, St Olaf College, Northfield, Minnesota 55057, USA. <sup>9</sup>Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996, USA.



# Haematopoietic stem cell induction by somite-derived endothelial cells controlled by *meox1*

Phong Dang Nguyen<sup>1\*</sup>, Georgina Elizabeth Hollway<sup>2,3\*</sup>, Carmen Sonntag<sup>1</sup>, Lee Barry Miles<sup>1</sup>, Thomas Edward Hall<sup>1</sup>, Silke Berger<sup>1</sup>, Kristine Joy Fernandez<sup>2</sup>, David Baruch Gurevich<sup>1</sup>, Nicholas James Cole<sup>4,5</sup>, Sara Alaei<sup>1,6</sup>, Mirana Ramialison<sup>1</sup>, Robert Lyndsay Sutherland<sup>2,3,†</sup>, Jose Maria Polo<sup>1,6</sup>, Graham John Lieschke<sup>1</sup> & Peter David Currie<sup>1,7\*</sup>

**Haematopoietic stem cells (HSCs) are self-renewing stem cells capable of replenishing all blood lineages. In all vertebrate embryos that have been studied, definitive HSCs are generated initially within the dorsal aorta (DA) of the embryonic vasculature by a series of poorly understood inductive events<sup>1–3</sup>. Previous studies have identified that signalling relayed from adjacent somites coordinates HSC induction, but the nature of this signal has remained elusive<sup>4</sup>. Here we reveal that somite specification of HSCs occurs via the deployment of a specific endothelial precursor population, which arises within a sub-compartment of the zebrafish somite that we have defined as the endotome. Endothelial cells of the endotome are specified within the nascent somite by the activity of the homeobox gene *meox1*. Specified endotomal cells consequently migrate and colonize the DA, where they induce HSC formation through the deployment of chemokine signalling activated in these cells during endotome formation. Loss of *meox1* activity expands the endotome at the expense of a second somitic cell type, the muscle precursors of the dermomyotomal equivalent in zebrafish, the external cell layer. The resulting increase in endotome-derived cells that migrate to colonize the DA generates a dramatic increase in chemokine-dependent HSC induction. This study reveals the molecular basis for a novel somite lineage restriction mechanism and defines a new paradigm in induction of definitive HSCs.**

Cell lineage studies have demonstrated two cellular sources for the amniote DA, the splanchnic and somitic mesoderm<sup>5–12</sup>. The zebrafish DA, by contrast, is believed to originate *in toto* from cells of the lateral plate mesoderm, an assumption based on imaging using lateral plate mesoderm-specific transgenes<sup>13–18</sup>. Consequently, these analyses do not exclude a somitic origin for vasculature progenitors in zebrafish. Our analyses of the *choker* (*cho*) mutation<sup>19</sup> revealed defects in somite lineages. Secondary trunk myogenesis is reduced in *cho* mutants (Fig. 1a–h, k, l, Extended Data Fig. 1i–l, o), as are appendicular and hypaxial muscles and their progenitors (Fig. 1i–n, Extended Data Fig. 1a–h, m, n, Supplementary Videos 1 and 2). These cell types derive from the external cell layer (ECL)<sup>20,21</sup> and ECL cell numbers are reduced in *cho* homozygotes (Fig. 1o–s). The *cho* mutation results from a null mutation in the *meox1* gene (Extended Data Fig. 2a–h) and *meox1* is expressed within the early somite, becomes restricted to the ECL and consequently to appendicular muscle populations, which are disrupted in *cho* mutants (Fig. 1t–z, Extended Data Fig. 2i–n). *meox1* is also expressed in vascular-associated cells (VACs, arrowheads Fig. 1x'', y, z, Extended Data Fig. 2m) located adjacent to the DA.

DA marker expression was variably expanded in *meox1* mutants (Fig. 2a–j, Extended Data Fig. 3a–j). The most informative of these markers was *cxcl12b* (also known as *sdf1b*), which is localized initially

to a central region of newly formed somites and then in cells positioned in a salt and pepper manner within the DA<sup>22</sup> (Fig. 2a–j). Somite and DA-associated expression of *cxcl12b* is expanded in *meox1* mutants ( $n = 57$ ) with the global DA marker genes *notch3* ( $n = 16$ ) and *ephB2* ( $n = 13$ )<sup>23</sup> more modestly expanded (Extended Data Fig. 3i, j, and data not shown). Furthermore, *pax3a* GFP transgenic embryos, which express green fluorescent protein (GFP) within the early somite, exhibit perduring GFP within the ECL and endothelial cells (ECs) of the DA (Fig. 2k–p, Extended Data Fig. 3m–o', v–x'), with the majority of the *pax3a*-GFP-positive cells expressing *cxcl12b* (Fig. 2n–n', arrows). Furthermore, anti-Meox1 antibody staining reveals DA cells contain perduring Meox1 protein in a similar pattern to *cxcl12b*, despite the DA not being a site of *meox1* mRNA expression (Fig. 2q–q'', Extended Data Fig. 3k–l'''). Collectively, these studies suggest a somitic origin for a portion of cells of the zebrafish DA.

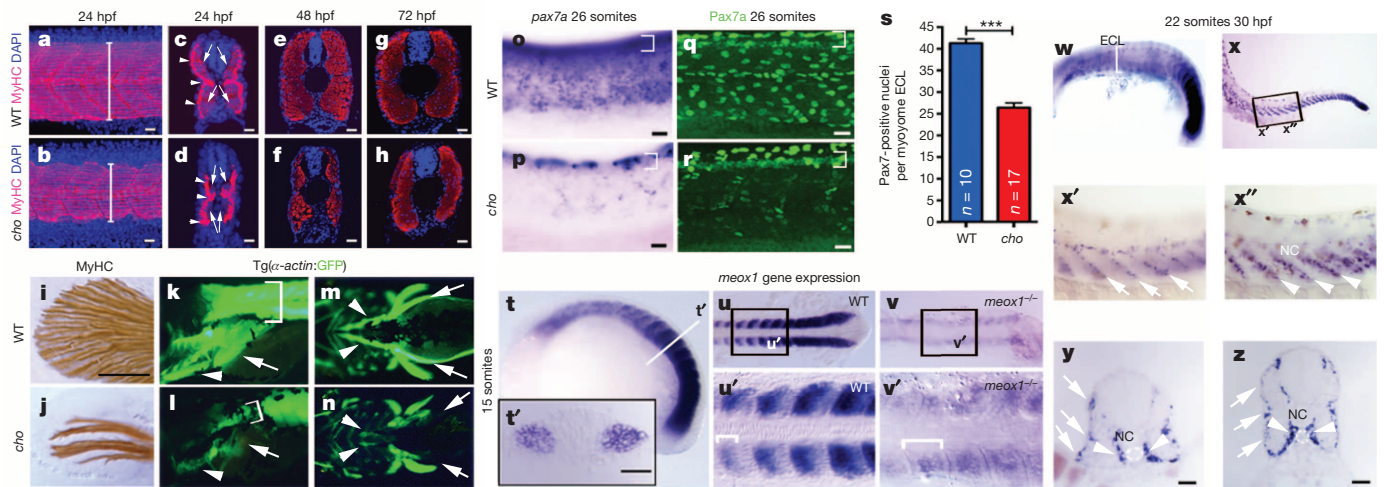
To test this hypothesis, somitic cell fate was analysed using the Kaede photoconvertible protein expressed globally ( $n = 23$ ) or from the somite-specific mesogenin (*msgn*) promoter<sup>24</sup> ( $n = 29$ , Fig. 3a–f'', Supplementary Videos 3 and 4, Extended Data Fig. 3p–u, Extended data Fig. 4). This analysis revealed that the anterior somite gives rise to the ECL and its derivatives as previously documented<sup>20,21</sup>, but also contributes to the DA, posterior cardinal vein (PCV) and inter-segmental blood vessels (ISVs), (Fig. 3a–d'', Extended data Fig. 3p–u, Supplementary Video 5, Extended Data Fig. 4b–h) and VACs (Fig. 3d'', arrows). Single-cell lineage analysis of the entire somite (Fig. 3h–i''', o,  $n = 96$ ) revealed that cells residing at the anterior/posterior midpoint of the somite provide progenitors to both the vasculature (Fig. 3o) and VACs (Fig. 3i–i'''). Furthermore, a single-cell Cre/loxP fate mapping technique was developed that demonstrated that central somitic cells are unipotent progenitors that undergo little cell division in contributing to DA formation ( $n = 7$ , Fig. 3j–j'', n). Next we generated a transgenic line that expressed tamoxifen-inducible Cre from the somite-specific *msgn1* promoter Tg(*msgn1*:CreERT2). Crossing this line to Tg(*βactin2*: loxP AcGFP1-STOP pA loxP mCherry pA; *βactin2* is also known as *actb2*) transgenics resulted in DA ECs expressing mCherry upon tamoxifen addition at early somite stages ( $n = 5$ , Fig. 3g–g''). Thus, four independent fate mapping strategies have demonstrated that DA cells can derive from a central somitic region we term the 'endotome'.

Fate mapping further revealed that *meox1* mutants (74%,  $n = 14$ , Fig. 3k, k', m–m'', o) but not their wild-type siblings (0%,  $n = 45$ ) can generate DA or VACs from anterior-most somite cells. Reciprocally, labelling of the anterior-most somite cells in *meox1* mutants resulted in fewer ECL cells and their derivatives (26%,  $n = 5$ ). However, the generation of primary muscle is unaffected in *meox1* mutants ( $n = 5$ , Fig. 3l, l', o) and loss of *meox1* did not alter the proliferation of somite cells either globally or in a lineage-specific manner (Extended Data Fig. 6),

<sup>1</sup>Australian Regenerative Medicine Institute, Level 1, Building 75, Monash University, Wellington Road, Clayton, Victoria 3800, Australia. <sup>2</sup>The Kinghorn Cancer Centre & Cancer Research Program, Garvan Institute of Medical Research, Victoria Street, Darlinghurst, New South Wales 2010, Australia. <sup>3</sup>St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Kensington, New South Wales 2052, Australia. <sup>4</sup>Anatomy Department, University of Sydney Medical School Anderson Stuart Building University of Sydney, Camperdown, New South Wales 2006, Australia. <sup>5</sup>MND Research Program, Australian School of Advanced Medicine, Macquarie University, New South Wales 2109, Australia. <sup>6</sup>Department of Anatomy and Developmental Biology, Monash University, Wellington Road, Clayton Victoria 3800, Australia. <sup>7</sup>EMBL Australia, Level 1, Building 75, Monash University, Wellington Road, Clayton, Victoria 3800, Australia.

\*These authors contributed equally to this work.

†Deceased.



**Figure 1** | *cho* mutants possess defects in the ECL and the muscles derived from it. **a–h**, Anti-myosin heavy chain (MyHC) staining reveals reduced muscle (brackets, **a, b**) in *cho* mutants. Arrowheads (**c, d**) slow muscle, arrows (**c, d**) fast muscle. **i–n**, Reduced trunk (100%,  $n = 60$ , brackets) fin (66%,  $n = 32$ ), hypaxial (65%,  $n = 60$ , arrows) and sternohyoides (100%  $n = 60$ , arrowheads) muscle in *cho* mutants. Lateral (**k, l**) and ventral (**m, n**) views of 72 hours post fertilisation (hpf) Tg( $\alpha$ -actin:GFP) larvae. **o–s**, *pax7a* mRNA (**o, p**) and protein expressing cells (**q, r, s**) are reduced in number within the ECL but normal in the dorsal neural tube (bracket). **s**, ECL quantitation:

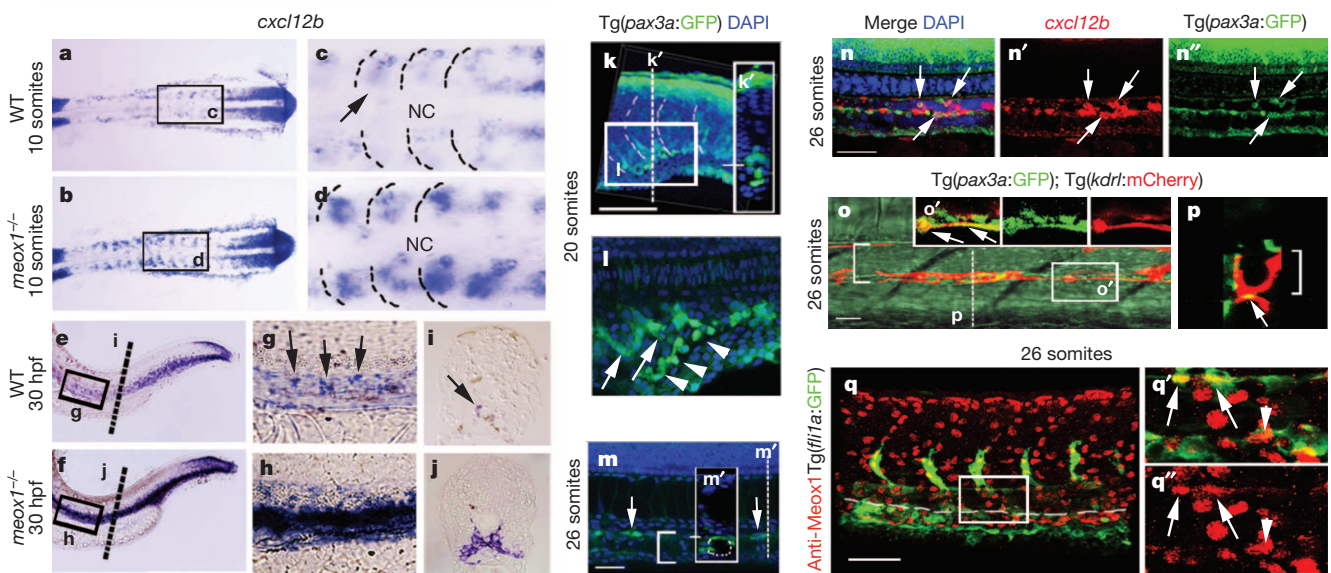
mean  $\pm$  s.e.m.; significance (\*\*\*)  $P < 0.0001$  in unpaired *t*-test. **t–v**, *meox1* expression 15 somites. **t'**, Cross-section as in **t**. **u'**, area boxed in **u** (bracket). **v, v'**, *meox1* transcript levels suggests nonsense mediated decay in *cho* mutants. **v'**, region boxed in **v**. *cho* mutants lose anterior restriction of remaining *meox1* transcripts (compare brackets **u'** and **v'**). **w**, 22 somites, *meox1* is ECL localized. **x–z**, 30 hpf. *meox1* in the ECL, and myosepta of tail somites (arrows, **x'**, **x''**). **y, z**, cross-section within pre-yolk (**y**) and post-yolk (**z**) extension somites. Arrows, ECL; arrowheads, VACs; dashed circle, dorsal aorta. Scale bars, 20  $\mu$ m.

but did result in a transitory increase in the size of the PCV (Extended Data Fig. 5). Thus, *meox1* is required to partition the fate of the somite and loss of *meox1* expands the endotome at the expense of ECL progenitors. Reciprocally, overexpression of *meox1* induces expression of the ECL marker Pax7 in cells destined to be ECs (Extended Data Fig. 7).

Studies in chick embryos have suggested that the colonization of the DA with somite-derived ECs could negatively regulate the haematopoietic competence of the DA, as the haemogenic endothelium of the DA is replaced entirely by somite-derived ECs during embryogenesis<sup>9,10</sup>. We therefore examined whether expression of HSC markers was altered in

*meox1* mutants. This analysis surprisingly revealed an expansion of HSC marker expression, rather than the reduction predicted by an expanded colonization of haematopoietically refractory, somite-derived ECs in the DA of *meox1* mutants.

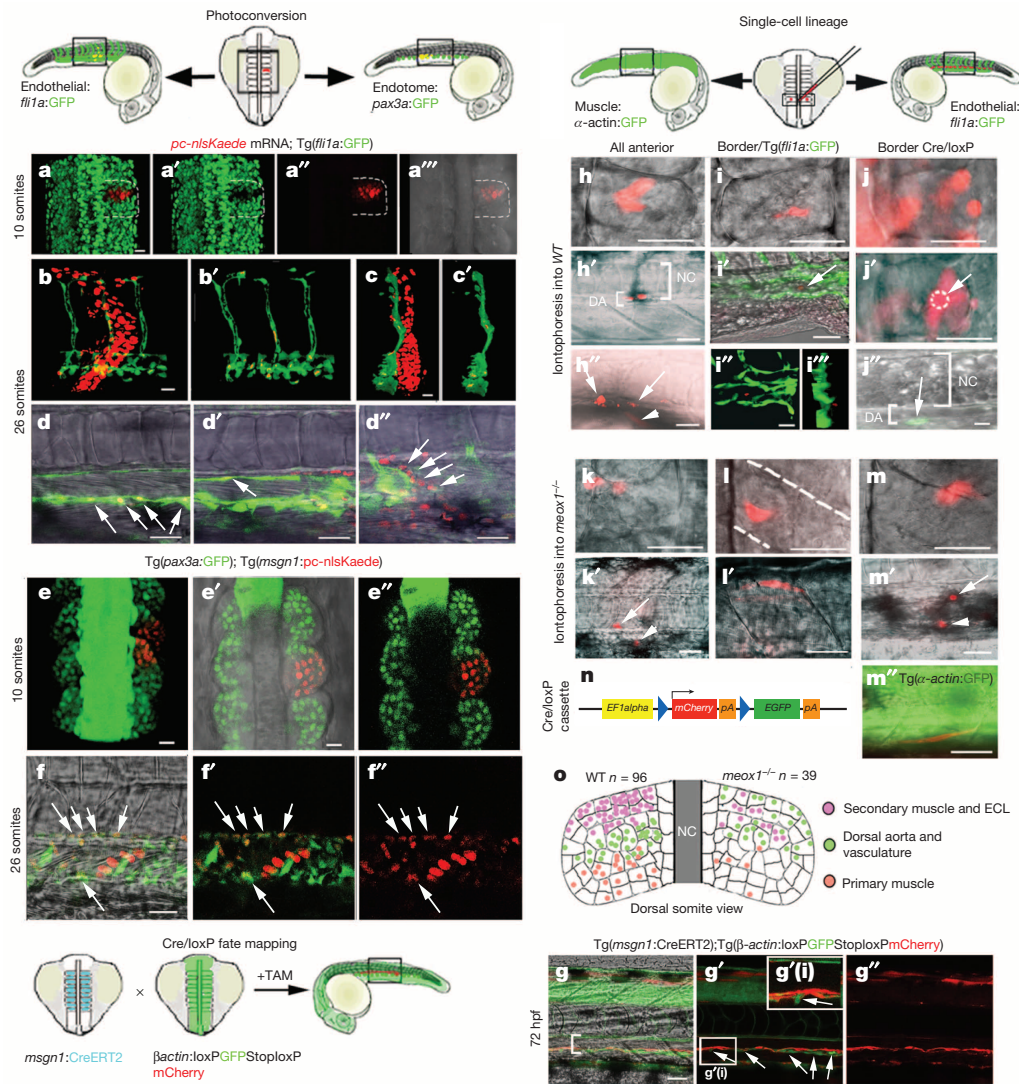
Specifically, the *itga2b*:GFP transgenic line (also known as *cd41*:GFP), which drives low-level GFP expression in DA-associated HSCs, and the genes *scl* (also known as *tal1*), *cmyb* and *runx1*, which mark HSCs in the DA, all showed expanded expression in *meox1* mutants or morphants (*scl*  $n = 42$ , Fig. 4a–h; *itga2b*:GFP, Extended Data Fig. 8a–i'; *runx1*  $n = 26$ , Extended Data Fig. 8j–o; *cmyb*  $n = 32$ , Extended Data Fig. 8p–u).



**Figure 2** | Endotome markers persist in the DA and are expanded in *meox1* mutants. *cxcl12b* is restricted to the mediolateral somite in wild type (WT) (**a, c**) but expands into the anterior somite in *meox1*<sup>-/-</sup> embryos (**b, d**) at 10 somites. Salt and pepper *cxcl12b* expression in the DA (arrows in **g** and **i**) is expanded in *meox1*<sup>-/-</sup> embryos (**f, h, j**) at 30 hpf. **a–d**, dorsal views; **e–h**, lateral views; **i, j**, cross-section. **k, l**, Tg(*pax3a*:GFP) (green) marks ventral (arrows, **l**) and migrating (arrowheads, **l**) somitic cells. DAPI, blue. **k'**, Transverse section as in **k**. **l**, Confocal section of the region boxed in **k**.

**m**, 26 somites, GFP in DA cells (arrows). Confocal section at mid-DA. **m'**, Transverse section as in **m**. **n–n''**, Confocal section mid-DA reveals *pax3a*-GFP cells express *cxcl12b* (red, arrow). **o–p**, *pax3a*-GFP cells co-express Tg(*kdr*:mCherry) (arrow, DA brackets). **o'**, Area boxed in **o**. **p**, Transverse section as in **o**. **q–q''**, Anti-Meox1 (red) and anti-GFP (green) co-localize in the dorsal (arrows, **q'**, **q''**) and ventral (arrowheads, **q'**–**q''**) DA. Scale bars, 50  $\mu$ m; **o**, 20  $\mu$ m.



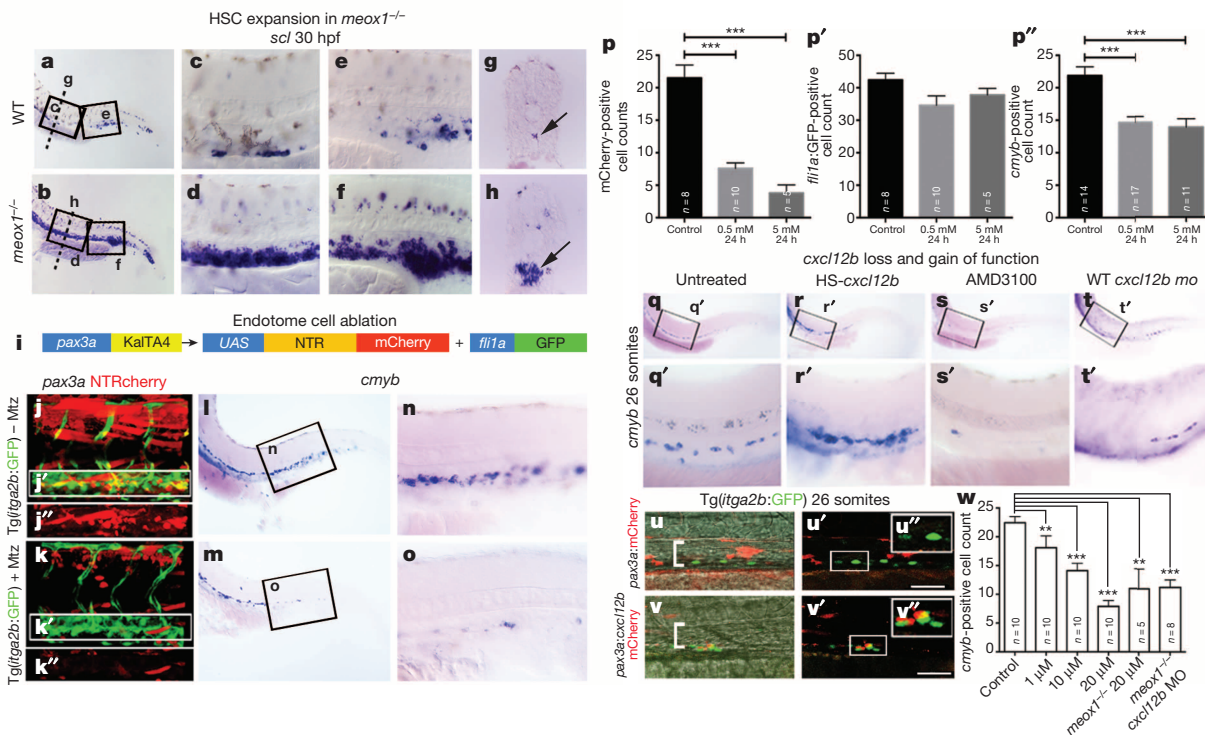


**Figure 3 | Endotome cells contribute to DA formation and are expanded in *meox1* mutants.** a–f', a–a'', nlsKaede (green) photoconverted (red) in the anterior somite ( $n = 23$ , 10 somites). b–d', Same embryo at 26 somites in *fli1a*-GFP. b–b', *fli1a*-GFP (green) and photoconverted endotome cells. c, c', Segmentation for vasculature colocalization (yellow). Scale Bar 20 μm. d–d'', Endotome-derived nuclei (arrows) in ECs (green) of ventral (d) and dorsal DA (d') and VACs (d'). e–f', 10 somites TgBAC(*pax3a*:GFP) (green), Tg(*msgn1*:nlsKaede) photoconverted embryos (e–e'). f–f'', At 26 somites photoconverted nuclei co-localize with GFP in Tg(*pax3a*:GFP) (100%,  $n = 29$ , arrows). g–g'', Tamoxifen added at 10 somites in Tg(*msgn1*:CreERT2); Tg(*βactin2*: loxP AcGFP1-STOP pA loxP mCherry pA) animals results in mCherry in DA cells ( $n = 5$ , bracket) at 72 hpf. g' (i) region boxed in g'. h–h'', Anterior somite tetramethylrhodamine dextran (TMRD)-labelled cells at 10 somites (h, RD, red) localize within the DA (h'), ECL (h', arrows) and

We hypothesized that three mechanisms could generate the increase of HSCs evident in *meox1* mutants. First, endotome cells could directly contribute to HSC formation and the excess HSCs evident in *meox1* mutants could derive from increased endotome cell numbers. However, time-lapse and fate-mapping analyses failed to detect direct contribution of somite-derived cells to HSC formation in either wild-type ( $n = 23$ ) or *meox1*-deficient contexts ( $n = 35$ ), (Extended Data Fig. 9a–n'). Furthermore, somite cells marked using the *msgn1*:CreERT2 line failed to label cells budding from the ventral floor of the DA into the PCV, the known position of HSC cells, despite being intimately associated with these cells ( $n = 5$ , Fig. 3g–g'', arrows). In line with this observation, the DA expression of the *pax3a* transgene occurs adjacent to HSCs, but does not co-localize (Extended Data Fig. 3v–x'). Therefore, we concluded

that somite-derived ECs do not directly contribute to HSC formation in wild-type or *meox1* mutant embryos. Second, we examined whether *meox1* could influence the expression of *wnt16*, which has been previously postulated to indirectly regulate HSC generation through an undescribed 'relay signal'<sup>4</sup>. However, rather than exhibiting the upregulation of *wnt16* expression that this model predicts, *wnt16* expression was severely downregulated in *meox1* mutants (Extended Data Fig. 9o–x). Thus, the expansion of HSC markers evident in *meox1* mutants does not occur through an upregulation of *wnt16*, which appears dispensable for HSC induction. Lastly, we examined if somite-derived ECs could induce resident lateral-mesoderm-derived ECs of the DA to become HSCs. In support of such a model, somite-derived ECs are found adjacent to HSC clusters in the





**Figure 4 | Somite-derived endothelial cells are required for HSC induction.** **a-h**, *scl* expression in 30 hpf wild-type (WT) (**a**, **c**, **e**, **g**) and *meox1<sup>-/-</sup>* ( $n = 42$ , **b**, **d**, **f**, **h**) embryos. **c**, **e**, Regions boxed in **a**, **d**, **f**, Regions boxed in **b**, **g**, Section as in **a**, **h**, Section as in **b**, **i**, *pax3a:KalTA4* injected into Tg(*uas:E1b:Ec.NfsB-mCherry*) (abbreviated to Tg(*uas:NTRmCherry*)) and Tg(*fl1a:GFP*) incubated without (–) (**j-j''**, **l**, **n**) or with (+) (**k-k''**, **m**, **o**) metronidazole (Mtz). Mtz addition ablates *pax3a:NTRmCherry* cells from the DA (**k-k''**, **m**, **o**) compared to controls (**j-j''**, **l**, **n**) and results in a reduction in *cmyb*+ve HSCs at 30 hpf (**m**, **o**, **p''**) compared to controls (**l**, **n**, **p''**).

DA (Extended Data Figs 3v–x', 9a–f) with individual somite-derived ECs associated with the generation of multiple HSCs (Extended Data Fig. 9g–n'). Furthermore, a small but significant increase in proliferation of HSCs occurs in the context of *meox1* loss of function (Extended Data Fig. 8w). To ablate endotome-derived ECs we injected a *pax3a:KalTA4* construct into a Tg(*uas:E1b:Eco.NfsB-mCherry*) fish line which results in the expression of the nitroreductase enzyme specifically within somite-derived ECs. Addition of metronidazole (Mtz) to these embryos ablated somite-derived ECs and led to a reduction in HSCs (Fig. 4i–p'').

Genetic deletion of the genes encoding *Cxcl12* and its receptor *Cxcr4* has highlighted multiple roles for *Cxcl12* in induction, maintenance, proliferation and mobilization of murine HSCs within the bone marrow<sup>25–27</sup> roles that could be provided embryonically by *Cxcl12b* expression in somite-derived ECs. To test this hypothesis we both genetically and chemically interfered with *Cxcl12b* signalling during HSC induction. Loss of *Cxcl12b* function resulted in a reduction of HSCs, a deficit that can be rescued by endotome-specific expression of *Cxcl12b* (Fig. 4q–q', s–t', w, Extended Data Fig. 10e) and reduced *Cxcl12b* signalling rescued the HSC expansion evident in *meox1* mutants (Fig. 4w, Extended Data Fig. 10a–c'). Reciprocally, overexpression of *Cxcl12b* via induction of the heat-shock promoter (Fig. 4q–r') or from the endotome-specific *pax3a* promoter ( $n = 25$ , Fig. 4u–v', Extended Data Fig. 10d–d') resulted in clumps of HSC cells irregularly positioned along the DA. Furthermore, injection of *pax3a:cxcl12b* mCherry into the *itga2b:GFP* line revealed that HSC clumps are specifically associated with *Cxcl12b*:mCherry expressing cells ( $n = 27$ , Fig. 4u–v').

To study the molecular basis of *meox1* regulation of *cxcl12b* signaling in more detail, we transfected *meox1* into human ECs, which severely suppressed the level of *Cxcl12* expression in these cells (Extended Data

**p-p''**, Quantification: mean  $\pm$  s.e.m. significance (\*\*\*)  $P < 0.0001$  in unpaired *t*-test. **q-r'**, Untreated (**q, q'**) and heat shocked (HS) ( $n = 25$ , **r, r'**) HS *cxcl12b* embryos stained for *cmv*β. **s, s'**, AMD3100 treatment (20 μM) reduces HSC numbers in a dose-dependent manner (**w, t, t'**, *cxcl12b* morpholino injection reduces HSC numbers **u-v''**, *pax3a:cxcl12b* mCherry fusion injected into the Tg(*itga2b*:GFP) ( $n = 27$ ) reveals HSCs (green) cluster with *cxcl12*mCherry-expressing cells (**v-v''**) but not mCherry alone cells (**u-u''**). **w**, mean  $\pm$  s.e.m.; significance (\*\*\*)  $P < 0.0001$ , \*\*  $P < 0.005$  from unpaired *t*-test. **r'-t'(i)**, Scale bars, 20 μm.

Fig. 10f). Furthermore, Meox1-mediated chromatin immunoprecipitation of the *cxcl12b* locus from zebrafish embryos indicates that this regulation is direct (Extended Data Fig. 10g, h).

Much emphasis has been placed on attempting to understand the cellular and molecular mechanisms underlying HSC induction, due to the obvious and far reaching implications this process holds for HSC derived therapies. The existence of somite-derived ECs in the DA of both chick and mouse embryos suggests that the mechanism we describe here could be reproduced during amniote HSC induction. We suggest that the difficulty in inducing HSCs from pluripotent cells *in vitro*<sup>28</sup> may in some part be due to a requirement for somitically-derived ECs to participate in this process and provide an embryonic niche for HSC induction (Extended Data Fig. 10i). An obvious imperative is to determine what molecular signals are produced by somite-derived ECs, apart from the ones we have defined here, to coordinate HSC induction, an understanding that could lead to the efficient generation of well-defined, and clinically applicable HSCs *in vitro* and *in vivo*.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 January; accepted 14 July 2014.

Published online 13 August 2014.

- Davidson, A. J. & Zon, L. I. The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. *Oncogene* **23**, 7233–7246 (2004).
- Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–115 (2010).
- Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108–111 (2010).

4. Clements, W. K. *et al.* A somitic Wnt16/Notch pathway specifies haematopoietic stem cells. *Nature* **474**, 220–224 (2011).
5. Noden, D. M. Embryonic origins and assembly of blood vessels. *Am. Rev. Respir. Dis.* **140**, 1097–1103 (1989).
6. Pardanaud, L. & Dieterlen-Lievre, F. Does the paraxial mesoderm of the avian embryo have hemangioblastic capacity? *Anat. Embryol. (Berl.)* **192**, 301–308 (1995).
7. Pardanaud, L. *et al.* Two distinct endothelial lineages in ontogeny, one of them related to hemopoiesis. *Development* **122**, 1363–1371 (1996).
8. Witting, J. *et al.* Angiogenic potential of the avian somite. *Dev. Dyn.* **202**, 165–171 (1995).
9. Jaffredo, T., Gautier, R., Eichmann, A. & Dieterlen-Lievre, F. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development* **125**, 4575–4583 (1998).
10. Pouget, C., Gautier, R., Teillet, M. A. & Jaffredo, T. Somite-derived cells replace ventral aortic hemangioblasts and provide aortic smooth muscle cells of the trunk. *Development* **133**, 1013–1022 (2006).
11. Esner, M. *et al.* Smooth muscle of the dorsal aorta shares a common clonal origin with skeletal muscle of the myotome. *Development* **133**, 737–749 (2006).
12. Ben-Yair, R. & Kalcheim, C. Notch and bone morphogenetic protein differentially act on dermomyotome cells to generate endothelium, smooth, and striated muscle. *J. Cell Biol.* **180**, 607–618 (2008).
13. Childs, S., Chen, J. N., Garrity, D. M. & Fishman, M. C. Patterning of angiogenesis in the zebrafish embryo. *Development* **129**, 973–982 (2002).
14. Jin, S. W., Beis, D., Mitchell, T., Chen, J. N. & Stainier, D. Y. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199–5209 (2005).
15. Kohli, V., Schumacher, J. A., Desai, S. P., Rehn, K. & Sumanas, S. Arterial and venous progenitors of the major axial vessels originate at distinct locations. *Dev. Cell* **25**, 196–206 (2013).
16. Lawson, N. D. & Weinstein, B. M. *In vivo* imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* **248**, 307–318 (2002).
17. Williams, C. *et al.* Hedgehog signaling induces arterial endothelial cell formation by repressing venous cell fate. *Dev. Biol.* **341**, 196–204 (2010).
18. Zhang, X. Y. & Rodaway, A. R. SCL-GFP transgenic zebrafish: *in vivo* imaging of blood and endothelial development and identification of the initial site of definitive hematopoiesis. *Dev. Biol.* **307**, 179–194 (2007).
19. Svetic, V. *et al.* Sdf1a patterns zebrafish melanophores and links the somite and melanophore pattern defects in choker mutants. *Development* **134**, 1011–1022 (2007).
20. Hollway, G. E. *et al.* Whole-somite rotation generates muscle progenitor cell compartments in the developing zebrafish embryo. *Dev. Cell* **12**, 207–219 (2007).
21. Stellabotte, F., Dobbs-McAuliffe, B., Fernandez, D. A., Feng, X. & Devoto, S. H. Dynamic somite cell rearrangements lead to distinct waves of myotome growth. *Development* **134**, 1253–1257 (2007).
22. Cha, Y. R. *et al.* Chemokine signaling directs trunk lymphatic network formation along the preexisting blood vasculature. *Dev. Cell* **22**, 824–836 (2012).
23. Bussmann, J., Bakkers, J. & Schulte-Merker, S. Early endocardial morphogenesis requires Scl/Tal1. *PLoS Genet.* **3**, e140 (2007).
24. Yabe, T. & Takada, S. Mesogenin causes embryonic mesoderm progenitors to differentiate during development of zebrafish tail somites. *Dev. Biol.* **370**, 213–222 (2012).
25. Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
26. Greenbaum, A. *et al.* CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* **495**, 227–230 (2013).
27. Zou, Y. R., Kottmann, A. H., Kuroda, M., Taniuchi, I. & Littman, D. R. Function of the chemokine receptor CXCR4 in hematopoiesis and in cerebellar development. *Nature* **393**, 595–599 (1998).
28. Cerdan, C. & Bhatia, M. Novel roles for Notch, Wnt and Hedgehog in hematopoiesis derived from human pluripotent stem cells. *Int. J. Dev. Biol.* **54**, 955–963 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. Kardon and B. Hogan for reading and critique of the manuscript. We also thank C.-H. Wang, F. Ellett, V. Nikolova-Krstevski and Fishcore staff for technical assistance. This work was supported by a National Health and Medical Research Council of Australia (NHMRC) grant to P.D.C. and an Australian Research Council grant to P.D.C. and G.E.H.; G.E.H. was supported by a Cancer Institute NSW (CINSW) Career Development Fellowship, R.L.S. by the CINSW and RT Hall Foundation, P.D.N. by an Australian Postgraduate Award, G.J.L. by a NHMRC Senior Research Fellowship and P.D.C. by a NHMRC Principal Research Fellowship. The Australian Regenerative Medicine Institute is supported by funds from the State Government of Victoria and the Australian Federal Government.

**Author Contributions** P.D.C. (fate mapping, mutant analyses), P.D.N. (fate mapping, micro injections, mutant analyses, transgenic construct generation) and G.E.H. (mutant cloning, construct generation, *in situ* hybridization, micro injections, mutant analyses and cell transfections) designed and performed experiments; C.S. (*in situ* hybridizations, microinjection), L.M. (confocal analyses), T.E.H. (generated transgenic constructs), S.B. (histology), K.J.F. (*in situ* hybridization, cell transfections), D.B.G. (generated transgenic constructs) and S.A. (ChIP) performed experiments; N.J.C. and R.L.S. provided reagents; G.J.L., M.R. and J.M.P. provided reagents and assisted with revisions; P.D.C., P.D.N. and G.E.H. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D.C. ([peter.currie@monash.edu](mailto:peter.currie@monash.edu)).

# Jam1a–Jam2a interactions regulate haematopoietic stem cell fate through Notch signalling

Isao Kobayashi<sup>1</sup>, Jingjing Kobayashi-Sun<sup>1</sup>, Albert D. Kim<sup>1</sup>, Claire Pouget<sup>1</sup>, Naonobu Fujita<sup>2</sup>, Toshio Suda<sup>3</sup> & David Traver<sup>1,2</sup>

Notch signalling plays a key role in the generation of haematopoietic stem cells (HSCs) during vertebrate development<sup>1–3</sup> and requires intimate contact between signal-emitting and signal-receiving cells, although little is known regarding when, where and how these intercellular events occur. We previously reported that the somitic Notch ligands, Dlc and Dld, are essential for HSC specification<sup>4</sup>. It has remained unclear, however, how these somitic requirements are connected to the later emergence of HSCs from the dorsal aorta. Here we show in zebrafish that Notch signalling establishes HSC fate as their shared vascular precursors migrate across the ventral face of the somite and that junctional adhesion molecules (JAMs) mediate this required Notch signal transduction. HSC precursors express *jam1a* (also known as *fli1r*) and migrate axially across the ventral somite, where Jam2a and the Notch ligands Dlc and Dld are expressed. Despite no alteration in the expression of Notch ligand or receptor genes, loss of function of *jam1a* led to loss of Notch signalling and loss of HSCs. Enforced activation of Notch in shared vascular precursors rescued HSCs in *jam1a* or *jam2a* deficient embryos. Together, these results indicate that Jam1a–Jam2a interactions facilitate the transduction of requisite Notch signals from the somite to the precursors of HSCs, and that these events occur well before formation of the dorsal aorta.

JAM proteins belong to the immunoglobulin superfamily of cell adhesion molecules, comprised of three closely related members, JAM1 (also known as JAM-A or F11R), JAM2 (also known as JAM-B), and JAM3 (also known as JAM-C)<sup>5</sup>. It has been reported that Jam1 is expressed in both murine and zebrafish HSC fractions<sup>6,7</sup>, although its role in haematopoiesis remains unknown. In zebrafish, the *jam1* gene was tandemly duplicated on chromosome 5 to generate *jam1a* and *jam1b* (also known as *fli1rl*). The structure of Jam1a is similar to that of human JAM1, which is composed of two immunoglobulin-like domains, a transmembrane domain (TM), and a PDZ-binding domain (PBD), whereas Jam1b lacks the TM and PBD (Extended Data Fig. 1a–d). We therefore focused on Jam1a to determine its potential roles in HSC development. We first examined the expression of *jam1a* in zebrafish embryos. At 14 h post-fertilization (hpf), *jam1a* was expressed in bilateral stripes of posterior lateral mesoderm (PLM) (Extended Data Fig. 2a), which gives rise to both endothelial and haematopoietic lineages<sup>8</sup>. After 18 hpf, however, *jam1a* was no longer detected in endothelial cells (Extended Data Fig. 2b, c). We performed co-staining of *jam1a* with *fli1*, a marker of the vascular lineage. The expression domain of *fli1* overlapped with that of *jam1a* at 14 hpf (Extended Data Fig. 2d), indicating that PLM cells indeed express *jam1a* at this stage. We observed the downregulation of *jam1a* in purified green fluorescent protein (GFP)-labelled endothelial cells (*fli1:GFP*<sup>+</sup> cells) from 14 to 20 hpf (Extended Data Fig. 2e).

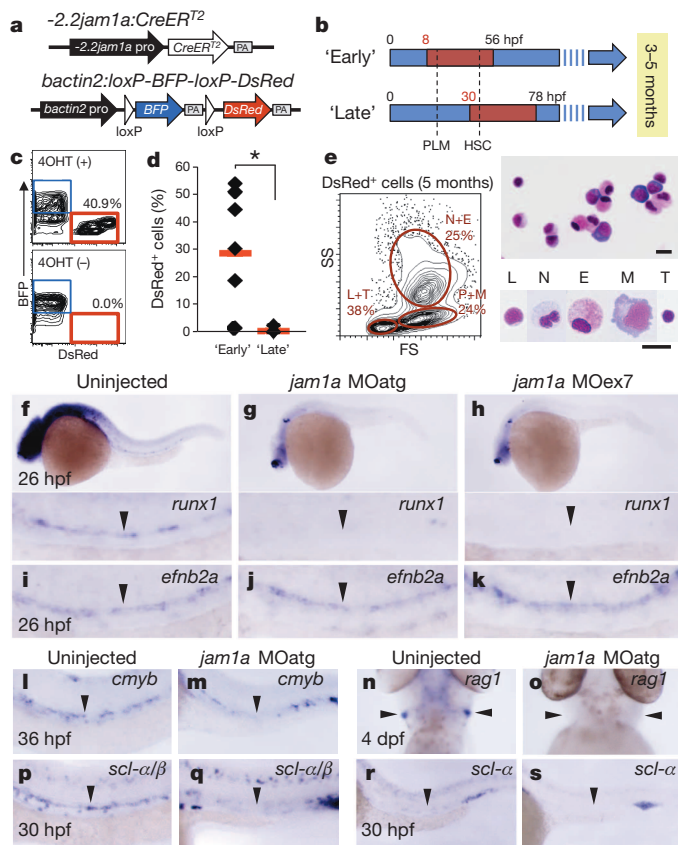
To determine if HSC precursors are contained within *jam1a*<sup>+</sup> PLM cells, we performed lineage tracing using the combined transgenic lines, *-2.2jam1a:CreER*<sup>T2</sup>, which expresses *CreER*<sup>T2</sup> under the control of *jam1a* regulatory elements, and *bactin2:loxP-BFP-loxP-DsRed*, which switches from expression of blue fluorescent protein (BFP) to the DsRed red fluorescent protein following Cre-based recombination (Fig. 1a, Extended Data Fig. 2f). Double-transgenic embryos were treated with

4-hydroxytamoxifen (4OHT) following two different schedules (Fig. 1b). An ‘early’ group was treated with 4OHT from 8 hpf, a stage before PLM formation<sup>9</sup>, and a ‘late’ group from 30 hpf, a stage just before HSC emergence in the dorsal aorta<sup>10,11</sup>. These embryos were grown to 3–5 months of age, after which whole kidney marrow cells were analysed by flow cytometry (Fig. 1c). As shown in Fig. 1d, high percentages of ‘switched’ DsRed<sup>+</sup> cells were detected in the ‘early’ group. DsRed<sup>+</sup> cells were comprised of multiple types of blood lineages (Fig. 1e). In contrast to the ‘early’ schedule, DsRed<sup>+</sup> cells were nearly undetectable in the ‘late’ group (Fig. 1d). These results indicate that *jam1a* is expressed in the shared vascular precursors of HSCs during early somitogenesis stages. The expression of *jam1a* in HSC precursors was further confirmed by additional lineage-tracing studies using a *-5.1jam1a:CreER*<sup>T2</sup> transgenic animal, which has an extended *jam1a* promoter/enhancer region (Extended Data Fig. 2g–l).

To examine the function of Jam1a in haematopoiesis, we designed two different morpholino oligonucleotides (MOs), *jam1a* MOatg (a translation-blocking MO) and MOex7 (a splice-blocking MO) (Extended Data Fig. 3a–e). We first examined the expression of the HSC marker gene *runx1* in these morphants. As shown in Fig. 1f, *runx1* was detected in the dorsal aorta in uninjected wild-type embryos at 26 hpf. In contrast, *runx1* was nearly undetectable in *jam1a* MOatg- and MOex7-injected embryos at the same stage (Fig. 1g, h). The expression of *efnb2a* (ephrin-B2a, a dorsal aorta marker gene) was unaffected in either morphant (Fig. 1i–k), suggesting that the dorsal aorta is specified normally. To further characterize *jam1a* morphants, we investigated the expression of additional marker genes. The expression of *cmyb* (another HSC marker) in the dorsal aorta was largely absent in *jam1a* morphants (Fig. 1l, m, Extended Data Fig. 3f, g). T-cell colonization of the thymus requires input from HSCs, providing a useful readout for whether HSCs have been specified or not. In *jam1a* morphants, the expression of *rag1* (a marker of immature T cells) was absent in the thymus at 4 days post-fertilization (dpf) (Fig. 1n, o, Extended Data Fig. 3h, i). A truncated isoform of *scl* (also known as *tall1*), *scl-β*, has been shown to mark haemogenic endothelium in the dorsal aorta<sup>12</sup>. Comparison of *scl-α/β* and *scl-α* probes revealed the specific reduction of *scl-β* in the dorsal aorta in *jam1a* morphants (Fig. 1p–s). Nascent HSCs can be visualized as *cmyb:GFP*; *kdr:mCherry* double-positive cells in the ventral floor of the dorsal aorta<sup>10</sup>. The number of double-positive cells in the dorsal aorta was twelve times lower in *jam1a* morphants than in wild-type embryos (Extended Data Fig. 4a–c). The expression of *gata1* (an erythroid marker) and *l-plastin* (a myeloid marker) at 24 hpf was normal in *jam1a* morphants, whereas the expression of *l-plastin* at 4 dpf was reduced in the caudal haematopoietic tissue (CHT) (Extended Data Fig. 4d–f). These results indicate that primitive haematopoiesis is unaffected, but definitive haematopoiesis is defective in *jam1a* morphants. The vasculature in the trunk was normal in *jam1a* morphants, whereas development of the vascular plexus in the CHT was slightly abnormal (Extended Data Fig. 4g–j). Development of the pronephros, somite, sclerotome and notochord was unaffected in *jam1a* morphants (Extended Data Fig. 4k–o). These results indicate that the failure of HSC specification in *jam1a* morphants is specific and not due to gross

<sup>1</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093-0380, USA. <sup>2</sup>Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, California 92093-0380, USA. <sup>3</sup>Department of Cell Differentiation, The Sakaguchi Laboratory, School of Medicine, Keio University, Shinjuku-ku, Tokyo 160-8582, Japan.





**Figure 1 | Loss of *jam1a* results in the loss of HSCs.** **a**, Vector constructs of transgenic animals used for lineage tracing. PA, polyA. **b**, Two different schedules of 4-hydroxytamoxifen (4OHT) treatment ('early' and 'late'). Red insets in the blue arrows indicate the period of the 4OHT treatment. **c**, Flow cytometric analysis of adult kidney marrow cells. **d**, The percentages of DsRed<sup>+</sup> cells in kidney marrow in the 'early' ( $n = 7$ ) or 'late' group ( $n = 10$ ). Red bars indicate the mean percentage. \* $P < 0.002$ , by Student's  $t$ -test. **e**, Flow cytometric and morphological analysis of DsRed<sup>+</sup> cells. L, lymphocytes, N, neutrophils; E, eosinophils; M, monocytes; T, thrombocytes, P, precursors. May-Grünwald-Giemsa staining. Bars, 10  $\mu$ m. **f–k**, Expression of *runx1* and *efnb2a* in uninjected, *jam1a* MOatg-, or MOex7-injected embryos. **l–s**, Expression of *cmyb*, *rag1*, *scl-α/β*, and *scl-α* in uninjected or *jam1a* MOatg-injected embryos. Arrowheads indicate the dorsal aorta (**f–m**, **p–s**) or thymus (**n**, **o**). Data are pooled from two independent experiments (**c–e**) or representative of two independent experiments with two different clutches of embryos (**f–s**).

malformations in adjacent environmental tissues. The effects of MOs are summarized in Supplementary Table 1.

Since *jam1a* is expressed in PLM cells, we next examined the formation and migration of the PLM in *jam1a* morphants. The expression of *fli1* at 12 hpf was normal in both types of *jam1a* morphants (Fig. 2a–c), suggesting that PLM formation is unaffected. PLM cells migrate axially and reach the midline by 17 hpf (Fig. 2d). We observed a delay in the migration of PLM cells in both types of *jam1a* morphants, in that a subset of *fli1*<sup>+</sup> cells did not reach the midline by 17 hpf (Fig. 2e, f). We performed time-lapse imaging of PLM cells from 14 hpf using *fli1:GFP*; *phldb1:mCherry* double transgenic embryos, where endothelial precursors and somitic cells are labelled by GFP and mCherry expression, respectively. PLM cells in the first wave reached the midline by 15.5 hpf in wild-type embryos, whereas the remaining cells reached the midline by 17.5 hpf to form the 'vascular cord' (Fig. 2g, Supplementary Video 1). In *jam1a* morphants, however, only a few PLM cells reached the midline by 15.5 hpf. Moreover, some PLM cells remained at the lateral borders of the somites at 17.5 hpf, and the vascular cord was discontinuous (Fig. 2h, Supplementary Videos 2 and 3). We examined the morphology of migrating PLM cells. In wild-type embryos, most migrating *fli1:GFP*<sup>+</sup> PLM cells displayed a flattened morphology

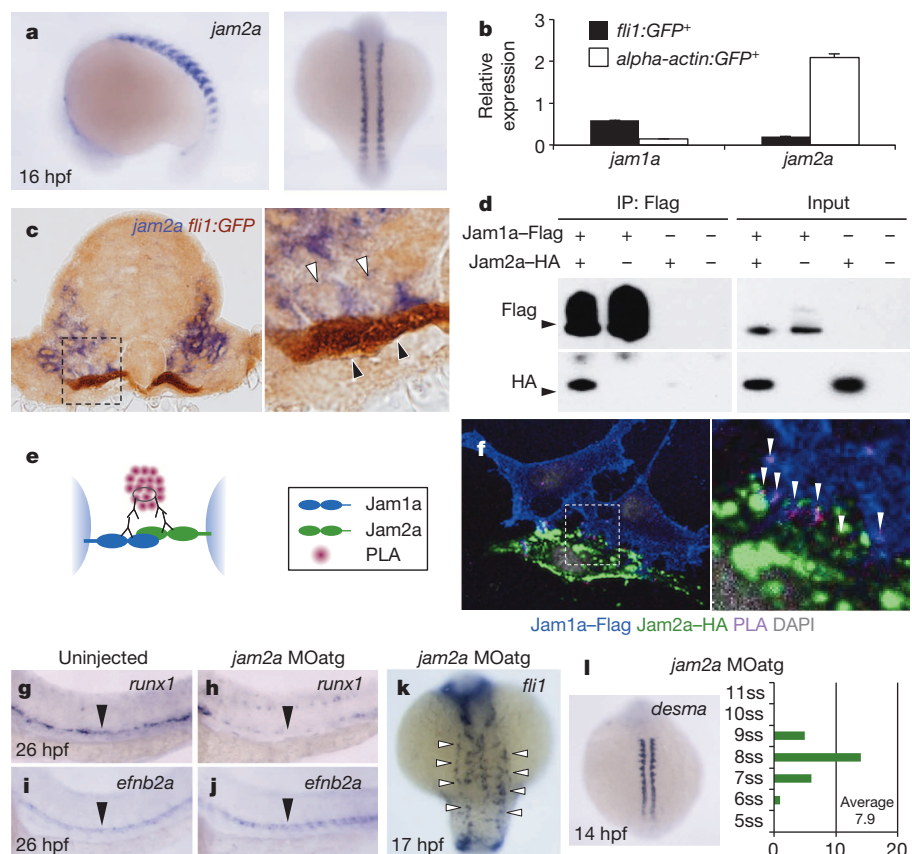
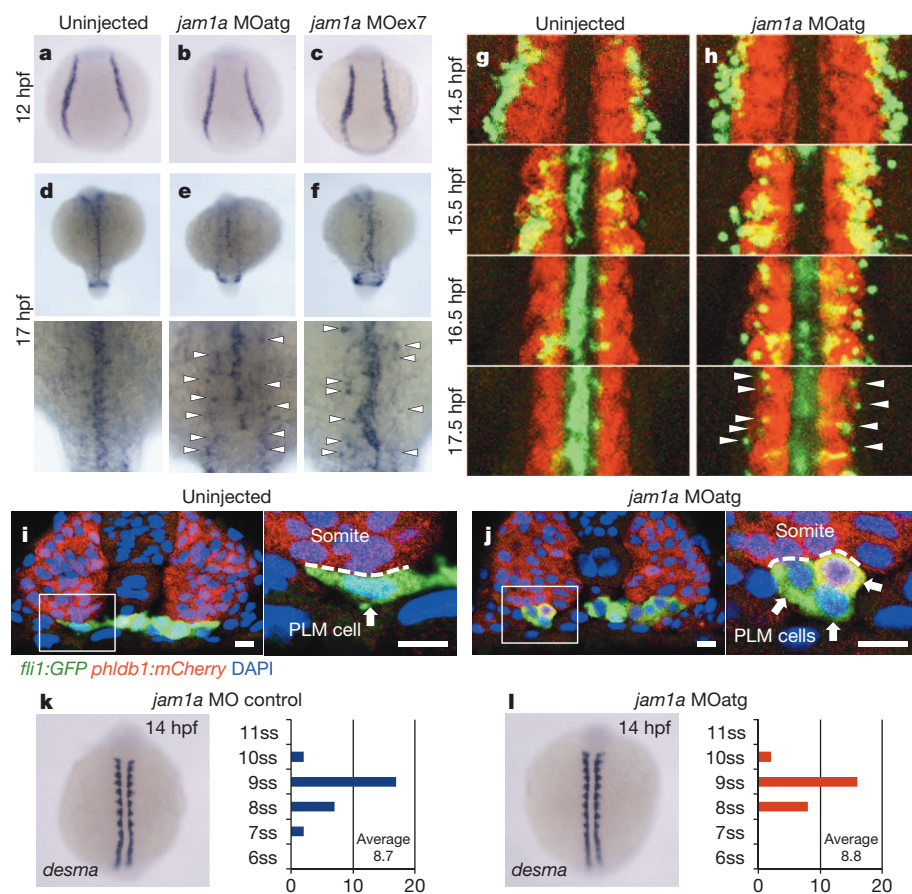
and appeared to interact closely with the ventral domain of the somite (Fig. 2i). By contrast, PLM cells in *jam1a* morphants displayed a round shape with relatively little surface contact with the somite (Fig. 2j).

To exclude the possibility of a general developmental delay in *jam1a* morphants, we enumerated somites at 14 hpf in *jam1a* MO control- and MOatg-injected embryos. We mainly observed nine somites formed in both *jam1a* MO control- and *jam1a* MOatg-injected embryos at this stage, and there was no significant difference in the average numbers of somites between groups (Fig. 2k, l). This indicates that the migration defect observed in *jam1a* morphants is specific and not due to developmental delay. In zebrafish, Hedgehog (Hh) and Vascular endothelial growth factor a (Vegfa) signalling pathways have been implicated to regulate the migration of PLM cells<sup>13,14</sup>. In *jam1a* morphants, however, the expression of *shha* (sonic hedgehog a) and *vegfa* as well as their downstream target *efnb2a* was unaffected (Fig. 1i–k, Extended Data Fig. 4n–q), indicating that the defect of PLM cell migration in *jam1a* morphants is independent of the Hh and Vegfa signalling pathways.

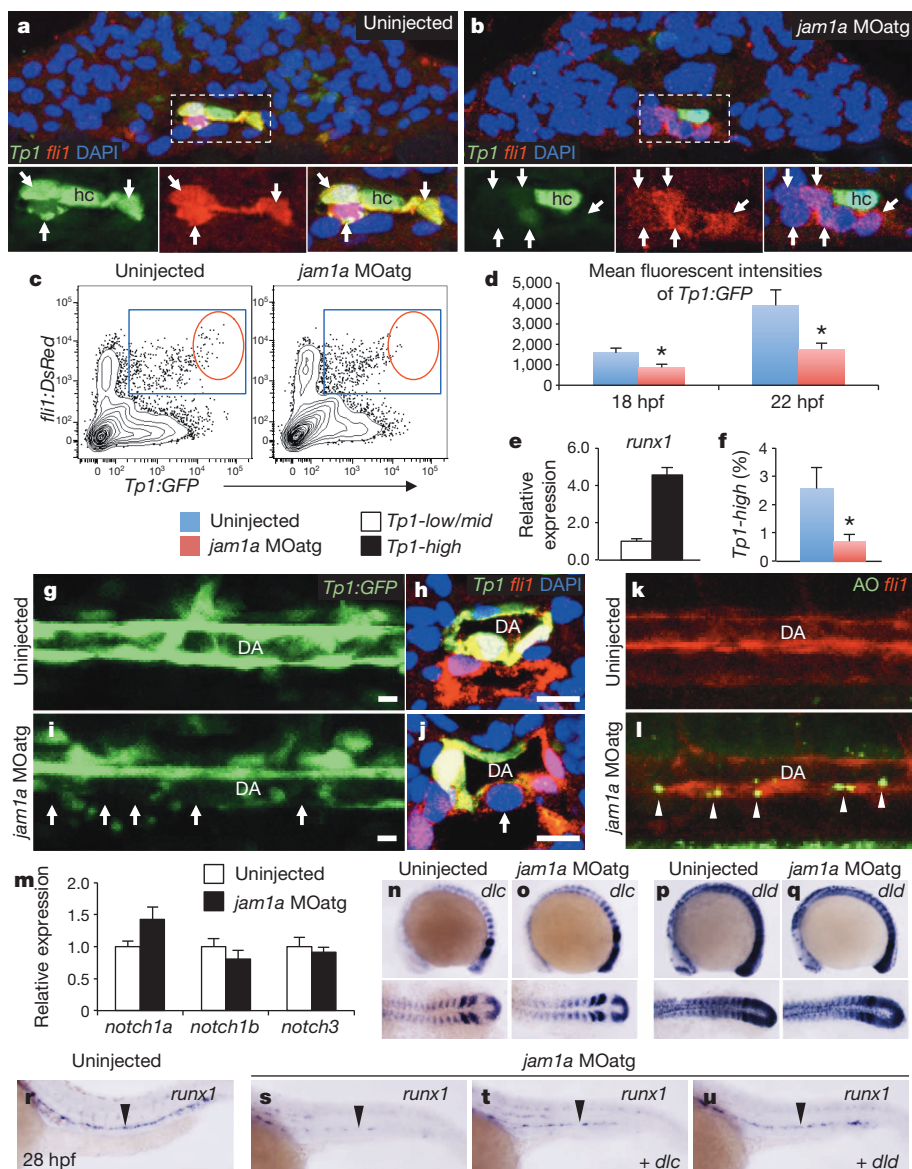
Because PLM cells migrate along the ventral domain of the somites, which includes the sclerotome, it is likely that a binding partner of Jam1a is expressed on the somitic epithelium. Previous studies determined the expression patterns of zebrafish *jam* genes and their physical binding properties by surface plasmon resonance. These studies showed that Jam1a can bind to Jam2a, Jam2b and Jam3a, but not to Jam1a (homotypically), Jam1b or Jam3b. Moreover, among these 6 *jam* genes, only *jam2a* and *jam3b* are expressed in somites<sup>15,16</sup>. Therefore, we next investigated whether PLM cells make contact with *jam2a*<sup>+</sup> somitic cells. As shown in Fig. 3a, *jam2a* was specifically expressed in somites at 16 hpf, a stage when PLM cells are migrating. Quantitative polymerase chain reaction (qPCR) results also showed that *jam2a* was highly expressed in purified *α-actin:GFP*<sup>+</sup> somitic cells at 14 hpf, whereas *jam1a* was highly expressed in purified *fli1:GFP*<sup>+</sup> PLM cells (Fig. 3b). Histological analysis of 16 hpf embryos revealed that migrating *fli1:GFP*<sup>+</sup> PLM cells were in close contact with *jam2a*<sup>+</sup> somitic cells (Fig. 3c).

To determine if Jam1a can bind to Jam2a, we performed coimmunoprecipitation experiments using transiently transfected Flag-tagged Jam1a (Jam1a-Flag) and haemagglutinin-tagged Jam2a (Jam2a-HA) constructs in HEK293T cells. Anti-Flag immunoprecipitation followed by anti-HA western blotting showed specific binding of Jam1a to Jam2a (Fig. 3d). To further test their interaction, we used a Duolink proximity ligation assay (PLA), which can demonstrate protein–protein interactions *in situ* by eliciting a fluorescent signal (Fig. 3e). As shown in Fig. 3f, PLA signals were detected in the boundary region between transfected Jam1a-Flag<sup>+</sup> cells and Jam2a-HA<sup>+</sup> cells, revealing the interaction of these proteins *in trans*. These results suggest that cells of PLM maintain intimate contact with cells of the ventral somite via Jam1a–Jam2a interactions during their migration.

This model predicts that loss of *jam2a* function would phenocopy the effects in *jam1a* morphants. We thus examined both HSC specification and PLM cell migration in *jam2a* MOatg- or MOex5-injected embryos (Extended Data Fig. 5a–d). The expression of *runx1* in the dorsal aorta was greatly reduced in both *jam2a* morphants, whereas *efnb2a* expression was unaffected (Fig. 3g–j, Extended Data Fig. 5e–l). In addition, axial migration of PLM cells was greatly delayed in both types of *jam2a* morphants (Fig. 3k, Supplementary Videos 4 and 5), despite only a modest delay in development (Fig. 3l). Migrating *fli1:GFP*<sup>+</sup> PLM cells in *jam2a* morphants displayed a round shape (Extended Data Fig. 5m), similar to that shown in *jam1a* morphants (Fig. 2j). The average contact surface area between a PLM cell and the somite was significantly reduced in both *jam1a* and *jam2a* morphants compared with uninjected embryos (Extended Data Fig. 5n). The effects of *jam2a* MOs were further validated in *jam2a* mutants (*jam2a*<sup>hu3319</sup>). Approximately 80% of homozygous *jam2a*<sup>hu3319</sup> embryos showed nearly undetectable expression of *runx1* and *cmyb* in the dorsal aorta and *rag1* in the thymus (Extended Data Fig. 6a–g). Moreover, approximately 85% of homozygous *jam2a*<sup>hu3319</sup> embryos showed delayed PLM cell migration compared with wild-type embryos (Extended Data Fig. 6h, i). Formation of the vasculature, however, was grossly normal







**Figure 4 | Notch signalling is depleted in *jam1a* morphants.** **a, b,** Transverse sections of *Tp1:GFP*; *fli1:DsRed* embryos un.injected or injected with *jam1a* MOatg at 18 hpf. Green and red channels and merges of the boxed regions are shown in the lower panels. Arrows indicate *fli1:DsRed*<sup>+</sup> cells. hc, hypochord. **c–f,** Flow cytometric and gene expression analysis of *Tp1:GFP*; *fli1:DsRed* embryos. Representative results of flow cytometric analysis at 22 hpf (**c**), the mean fluorescent intensities of GFP in *Tp1:GFP*<sup>+</sup>; *fli1:DsRed*<sup>+</sup> populations (**d**), relative expression levels of *runx1* in the *Tp1:GFP*<sup>high</sup> and *Tp1:GFP*<sup>low/mid</sup> population of *fli1:DsRed*<sup>+</sup> cells in wild-type embryos at 22 hpf (**e**), and the percentages of *Tp1:GFP*<sup>high</sup> in *fli1:DsRed*<sup>+</sup> populations at 22 hpf (**f**) are shown. Blue gates and red circles indicate the *Tp1:GFP*<sup>+</sup>; *fli1:DsRed*<sup>+</sup> and *Tp1:GFP*<sup>high</sup>; *fli1:DsRed*<sup>+</sup> population, respectively. \**P* < 0.01, by Student's *t*-test. Error bars, s.d. **g–j,** Lateral views of the dorsal aorta (DA) in *Tp1:GFP* embryos and transverse sections of *Tp1:GFP*; *fli1:DsRed* embryos un.injected or injected with *jam1a* MOatg at 28 hpf. Arrows indicate relatively low activation of *Tp1:GFP* in the ventral floor of the DA. Bars, 10  $\mu$ m. **k, l,** Acridine orange (AO) staining under the *fli1:DsRed* background in un.injected or *jam1a* MOatg-injected embryos at 30 hpf. Arrowheads indicate AO-stained apoptotic cells. **m,** Relative expression levels of *notch1a*, *notch1b*, and *notch3* in purified *fli1:GFP*<sup>+</sup> cells obtained from un.injected or *jam1a* MOatg-injected embryos at 18 hpf. Error bars, s.d. **n–q,** The expression of *dlc* and *dld* in un.injected or *jam1a* MOatg-injected embryos at 14 hpf. **r–u,** The expression of *runx1* at 26 hpf. Embryos were un.injected, injected with *jam1a* MOatg alone, or co-injected with *jam1a* MOatg and *dlc* or *dld* mRNA. Data are representative of two independent experiments with four embryos (**a, b, h, j**), eight embryos (**g, i, k, l**), four different clutches of embryos (**c–f**), or two different clutches of embryos (**m–u**).

in *jam2a*<sup>hu3319</sup> embryos (Extended Data Fig. 6j–o). These phenotypes are consistent with those in *jam1a* morphants, suggesting that Jam1a–Jam2a interactions are involved in both PLM cell migration and HSC specification.

Despite a large reduction in embryonic HSC number, approximately 50% of homozygous *jam2a*<sup>hu3319</sup> animals were viable and showed almost normal haematopoiesis in the adult kidney (Extended Data Fig. 6p, q). Further studies will be required to understand how haematopoiesis can recover in *jam2a*<sup>hu3319</sup> animals during development. Perhaps related to this observation, a dispensable role for Jam2 in adult haematopoiesis has also been reported in mice<sup>17,18</sup>.

To better understand how both *jam1a* and *jam2a* morphants show impaired HSC specification, we considered possible signal transduction mechanisms from the somite. Because our recent work demonstrated that two somitic Notch ligands, Dlc and Dld, are essential for HSC specification<sup>4</sup>, and because Notch is a juxtacrine signal that requires close contact between adjacent cells, we prioritized analysis of the Notch signalling pathway. To test the hypothesis that Jam1a–Jam2a interactions facilitate Notch signal transmission between the PLM and somite, we first examined the activation of Notch signalling in *jam1a* morphants using a Notch reporter line, *Tp1:GFP*, which expresses GFP under the control of tandem Notch responsive elements<sup>19</sup>. In wild-type embryos,

some *fli1:DsRed*<sup>+</sup> endothelial cells strongly expressed *Tp1:GFP* in the midline at 18 hpf (Fig. 4a). In *jam1a* morphants, by contrast, most of the *fli1:DsRed*<sup>+</sup> cells showed weak or no expression of the *Tp1:GFP* reporter at the same stage (Fig. 4b). The expression levels of *Tp1:GFP* in *fli1:DsRed*<sup>+</sup> cells were further quantified by flow cytometry (Fig. 4c). The mean fluorescence intensity of GFP in the *Tp1:GFP*<sup>+</sup>; *fli1:DsRed*<sup>+</sup> population was significantly lower in *jam1a* morphants than in un.injected embryos (Fig. 4d). In wild-type embryos, *runx1* is highly expressed in the *Tp1:GFP*<sup>high</sup> fraction of *fli1:DsRed*<sup>+</sup> cells at 22 hpf (Fig. 4e), suggesting that HSC precursors are enriched in this population. Notably, the percentage of *Tp1:GFP*<sup>high</sup>; *fli1:DsRed*<sup>+</sup> cells was significantly lower in *jam1a* morphants (Fig. 4f). At 28 hpf, *Tp1:GFP* was highly expressed in the dorsal aorta in wild-type embryos (Fig. 4g, h). Interestingly, in *jam1a* morphants, *Tp1:GFP* expression was weak and discontinuous along the floor of the dorsal aorta (Fig. 4i, j), the site of HSC emergence<sup>10,11</sup>. In addition, we observed many apoptotic cells along the aortic floor in *jam1a* morphants (Fig. 4k, l), suggesting that, in the absence of Notch signalling, HSC precursors fail to be specified and undergo apoptosis.

To test whether ectopic activation of Notch signalling is sufficient to rescue HSCs in *jam1a* morphants, we enforced expression of the Notch intracellular domain (NICD), a dominant activator of the Notch pathway<sup>3</sup>, using combined *hsp70:Gal4*; *UAS:NICD* transgenic lines. Heat-shock



induction of NICD at 14 hpf rescued the expression of *runx1* in the dorsal aorta in *jam1a* morphants (Extended Data Fig. 7a, b), similar to that shown previously for rescue of *mind bomb* (*mib*) mutants or *wnt16* morphants<sup>3,4</sup>. The expression of *runx1* was also rescued in *jam1a* morphants when NICD was induced in the PLM using the *fli1:Gal4* line (Extended Data Fig. 7c, d). Similar results were obtained in *jam2a* MOatg-injected embryos (Extended Data Fig. 7e, f).

At 15 hpf, migrating *fli1:GFP*<sup>+</sup> cells were observed to make direct contact with *dlc*<sup>+</sup> or *dld*<sup>+</sup> somitic cells (Extended Data Fig. 8a, b), indicating that PLM cells may receive Notch signalling via presentation of somitic Notch ligands. We observed low activation of *Tp1:GFP* in endothelial cells in *wnt16* morphants (Extended Data Fig. 8c–g), which show a reduction in somitic *dlc* and *dld*<sup>+</sup>. This suggests that Notch signalling in endothelial cells is activated at least in part by somitic Dlc and/or Dld. We investigated the expression of somitic Notch ligand genes (*dlc* and *dld*) as well as aortic Notch receptor and ligand genes (*notch1a*, *notch1b*, *notch3*, *dlc* and *delta-like 4* (*dll4*)) in *jam1a* morphants. Importantly, each was expressed normally in *jam1a* morphants (Fig. 4m–q, Extended Data Fig. 9a–h), suggesting that the defect in Notch signalling in *jam1a* morphants is due to low Notch signal transmission rather than to misregulation of Notch signalling components. Consistent with this postulate, we observed less contact surface area between migrating PLM cells and the somite in both *jam1a* and *jam2a* morphants (Fig. 2i, j, Extended Data Fig. 5m, n), which correlates with low activation of Notch signalling. Our hypothesis is further supported by an additional rescue experiment in which *dlc* or *dld* is globally overexpressed in *jam1a* morphants to present more Notch ligand to HSC precursors. As presented in Fig. 4u, the expression of *runx1* in the dorsal aorta was almost fully rescued by co-injection of *dld* messenger RNA (mRNA) along with the *jam1a* MOatg, whereas *runx1* expression was only partially rescued following co-injection with *dlc* mRNA (Fig. 4r–u). Furthermore, the expression of *Tp1:GFP* was also restored in the ventral floor of the dorsal aorta by co-injection with *dlc* or *dld* mRNA (Extended Data Fig. 9i–p). These data confirm that the impairment of HSC specification in *jam1a* morphants is caused by inadequate activation of Notch signalling in HSC precursors and suggest that Jam1a and Jam2a normally mediate the physical interaction between these precursors and the somite, which is required for efficient Notch signal transmission (Extended Data Fig. 10).

It has been reported that the overall levels of Notch signal transmission is proportional to adhesion strength between Notch receptor- and ligand-expressing cells<sup>20</sup>. Our data demonstrate that *runx1* is highly expressed in the *Tp1:GFP*<sup>high</sup> population of endothelial cells (Fig. 4e), suggesting that a relatively high level of Notch signalling is required to generate HSC fate. These findings strongly suggest that efficient Notch signal transduction in HSC precursors requires intimate intercellular contact mediated by Jam proteins. Moreover, our data suggest that HSC fate is established much earlier than previously appreciated, during the axial migration of PLM cells, which is well before formation of the dorsal aorta. These new findings may provide key insights into the timing and tissue interactions needed to instruct HSC fate, which should help inform *in vitro* approaches to generate HSCs from pluripotent stem cells.

## METHODS SUMMARY

For morpholino knockdown experiments, zygotes were injected with 1 nl of morpholino oligonucleotides (MOs, GeneTools). MO concentrations used were: *jam1a* MOatg (100  $\mu$ M), *jam1a* MO control (100  $\mu$ M), *jam1a* MOex7 (300  $\mu$ M), *jam2a* MOatg (300  $\mu$ M), *jam2a* MOex5 (400  $\mu$ M), and *wnt16* MO2 (5 ng nl<sup>-1</sup>). Fluorescent images were captured using an SP5 inverted confocal microscope (Leica) as previously described<sup>10</sup>. Flow cytometry, qPCR, *in situ* hybridization, histology, Duolink PLA, immunoprecipitation, and western blotting were performed as described in the Methods section.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 September 2013; accepted 27 June 2014.

Published online 13 August 2014.

- Hadland, B. K. *et al.* A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood* **104**, 3097–3105 (2004).
- Yoon, M. J. *et al.* Mind bomb-1 is essential for intraembryonic hematopoiesis in the aortic endothelium and the subaortic patches. *Mol. Cell. Biol.* **28**, 4794–4804 (2008).
- Burns, C. E., Traver, D., Mayhall, E., Shepard, J. L. & Zon, L. I. Hematopoietic stem cell fate is established by the Notch-Runx pathway. *Genes Dev.* **19**, 2331–2342 (2005).
- Clements, W. K. *et al.* A somitic Wnt16/Notch pathway specifies haematopoietic stem cells. *Nature* **474**, 220–224 (2011).
- Weber, C., Fraemohs, L. & Dejana, E. The role of junctional adhesion molecules in vascular inflammation. *Nature Rev. Immunol.* **7**, 467–477 (2007).
- Sugano, Y. *et al.* Junctional adhesion molecule-A, JAM-A, is a novel cell-surface marker for long-term repopulating hematopoietic stem cells. *Blood* **111**, 1167–1172 (2008).
- Kobayashi, I. *et al.* Comparative gene expression analysis of zebrafish and mammals identifies common regulators in hematopoietic stem cells. *Blood* **115**, e1–e9 (2010).
- Thompson, M. A. *et al.* The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. *Dev. Biol.* **197**, 248–269 (1998).
- Jin, S. W., Beis, D., Mitchell, T., Chen, J. N. & Stainier, D. Y. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199–5209 (2005).
- Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108–111 (2010).
- Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–115 (2010).
- Qian, F. *et al.* Distinct functions for different *scf* isoforms in zebrafish primitive and definitive hematopoiesis. *PLoS Biol.* **5**, e132 (2007).
- Gering, M. & Patient, R. Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. *Dev. Cell* **8**, 389–400 (2005).
- Wilkinson, R. N. *et al.* Hedgehog signaling via a calcitonin receptor-like receptor can induce arterial differentiation independently of VEGF signaling in zebrafish. *Blood* **120**, 477–488 (2012).
- Powell, G. T. & Wright, G. J. Jamb and Jamc are essential for vertebrate myocyte fusion. *PLoS Biol.* **9**, e1001216 (2011).
- Powell, G. T. & Wright, G. J. Genomic organisation, embryonic expression and biochemical interactions of the zebrafish junctional adhesion molecule family of receptors. *PLoS ONE* **7**, e40810 (2012).
- Sakaguchi, T. *et al.* Putative “stemness” gene jam-B is not required for maintenance of stem cell state in embryonic, neural, or hematopoietic stem cells. *Mol. Cell. Biol.* **26**, 6557–6570 (2006).
- Arcangeli, M. L. *et al.* JAM-B regulates maintenance of hematopoietic stem cells in the bone marrow. *Blood* **118**, 4609–4619 (2011).
- Parsons, M. J. *et al.* Notch-responsive cells initiate the secondary transition in larval zebrafish pancreas. *Mech. Dev.* **126**, 898–912 (2009).
- Ahimou, F., Mok, L. P., Bardot, B. & Wesley, C. The adhesion force of Notch with Delta and the rate of Notch signaling. *J. Cell Biol.* **167**, 1217–1229 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors thank G. Wright for providing the *jam2a*<sup>hu3319</sup> line, A. Shimizu for help in generating transgenic lines, M. Osato for providing the *I-SceI-pBSII SK+* vector, M. Distel for providing *phldb1:Gal4-mCherry* animals, and W. Clements, Y. Lee and E. Butko provided critical evaluation of the manuscript. This work was supported in part by a JSPS Research fellowship for Young Scientists and a JSPS Postdoctoral fellowship for Research Abroad (I.K.), by a New Investigator Award from the California Institute of Regenerative Medicine, R01-DK074482 from the National Institutes of Health, and an Innovative Science Award from the American Heart Association (D.T.).

**Author Contributions** I.K., T.S. and D.T. designed research. I.K. generated transgenic lines, performed flow cytometry, cell culture and transfection experiments, analysed data, and wrote the manuscript. I.K. and J.K.-S. performed *in situ* hybridization and real-time PCR. I.K., J.K.-S. and C.P. generated *in situ* probes. I.K. and N.F. performed immunoprecipitation and western blotting. I.K. and A.D.K. performed confocal imaging. J.K.-S. performed histological analyses. A.D.K., C.P., T.S. and D.T. edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.T. ([dtraver@ucsd.edu](mailto:dtraver@ucsd.edu)).

# A vaccine targeting mutant IDH1 induces antitumour immunity

Theresa Schumacher<sup>1,2\*</sup>, Lukas Bunse<sup>1,2\*</sup>, Stefan Pusch<sup>3,4</sup>, Felix Sahm<sup>3,4</sup>, Benedikt Wiestler<sup>1,5</sup>, Jasmin Quandt<sup>6</sup>, Oliver Menn<sup>1</sup>, Matthias Osswald<sup>1,5</sup>, Iris Oezen<sup>1,2</sup>, Martina Ott<sup>1,2</sup>, Melanie Keil<sup>1,2</sup>, Jörg Balß<sup>2,4</sup>, Katharina Rauschenbach<sup>1,2</sup>, Agnieszka K. Grabowska<sup>7</sup>, Isabel Vogler<sup>8</sup>, Jan Diekmann<sup>9</sup>, Nico Trautwein<sup>10</sup>, Stefan B. Eichmüller<sup>6</sup>, Jürgen Okun<sup>11</sup>, Stefan Stevanović<sup>10</sup>, Angelika B. Riemer<sup>7</sup>, Ugur Sahin<sup>9</sup>, Manuel A. Friese<sup>12</sup>, Philipp Beckhove<sup>6</sup>, Andreas von Deimling<sup>3,4</sup>, Wolfgang Wick<sup>1,5</sup> & Michael Platten<sup>1,2</sup>

**Monoallelic point mutations of isocitrate dehydrogenase type 1 (IDH1) are an early and defining event in the development of a subgroup of gliomas<sup>1–3</sup> and other types of tumour<sup>4–6</sup>. They almost uniformly occur in the critical arginine residue (Arg 132) in the catalytic pocket, resulting in a neomorphic enzymatic function, production of the oncometabolite 2-hydroxyglutarate (2-HG)<sup>7,8</sup>, genomic hypermethylation<sup>9–11</sup>, genetic instability and malignant transformation<sup>12</sup>. More than 70% of diffuse grade II and grade III gliomas carry the most frequent mutation, IDH1(R132H) (ref. 3). From an immunological perspective, IDH1(R132H) represents a potential target for immunotherapy as it is a tumour-specific potential neoantigen with high uniformity and penetrance expressed in all tumour cells<sup>13,14</sup>. Here we demonstrate that IDH1(R132H) contains an immunogenic epitope suitable for mutation-specific vaccination. Peptides encompassing the mutated region are presented on major histocompatibility complexes (MHC) class II and induce mutation-specific CD4<sup>+</sup> T-helper-1 (T<sub>H</sub>1) responses. CD4<sup>+</sup> T<sub>H</sub>1 cells and antibodies spontaneously occurring in patients with IDH1(R132H)-mutated gliomas specifically recognize IDH1(R132H). Peptide vaccination of mice devoid of mouse MHC and transgenic for human MHC class I and II with IDH1(R132H) p123–142 results in an effective MHC class II-restricted mutation-specific antitumour immune response and control of pre-established syngeneic IDH1(R132H)-expressing tumours in a CD4<sup>+</sup> T-cell-dependent manner. As IDH1(R132H) is present in all tumour cells of these slow-growing gliomas<sup>15</sup>, a mutation-specific anti-IDH1(R132H) vaccine may represent a viable novel therapeutic strategy for IDH1(R132H)-mutated tumours.**

To explore the immunogenicity of IDH1(R132H), 10-mer and 15-mer peptide libraries were generated encompassing the mutated (R132H) and the corresponding wild-type residue of human IDH1 (Extended Data Fig. 1a–c). Neither interrogation of MHC binding prediction algorithms nor T2 binding assays (Extended Data Fig. 1d–f) demonstrated peptide binding to the common MHC class I allele human leukocyte antigen (HLA)-A\*0201. In contrast, as shown by MHC class II binding assay, 15-mer peptides bound to the common MHC class II allele HLA-DRB1\*0101 (Fig. 1a and Extended Data Fig. 1g). Vaccination of MHC-humanized HLA-A\*0201 HLA-DRA\*0101 HLA-DRB1\*0101 transgenic mice devoid of mouse MHC (A2.DR1 mice) (ref. 16) with a p123–142 (R132H) peptide resulted in a robust interferon (IFN)- $\gamma$  T-cell response corresponding to the demonstrated peptide HLA-DRB1\*0101 binding strengths. Importantly, the T-cell response discriminated between mutated and wild-type IDH1 peptides (Fig. 1b, c and Extended

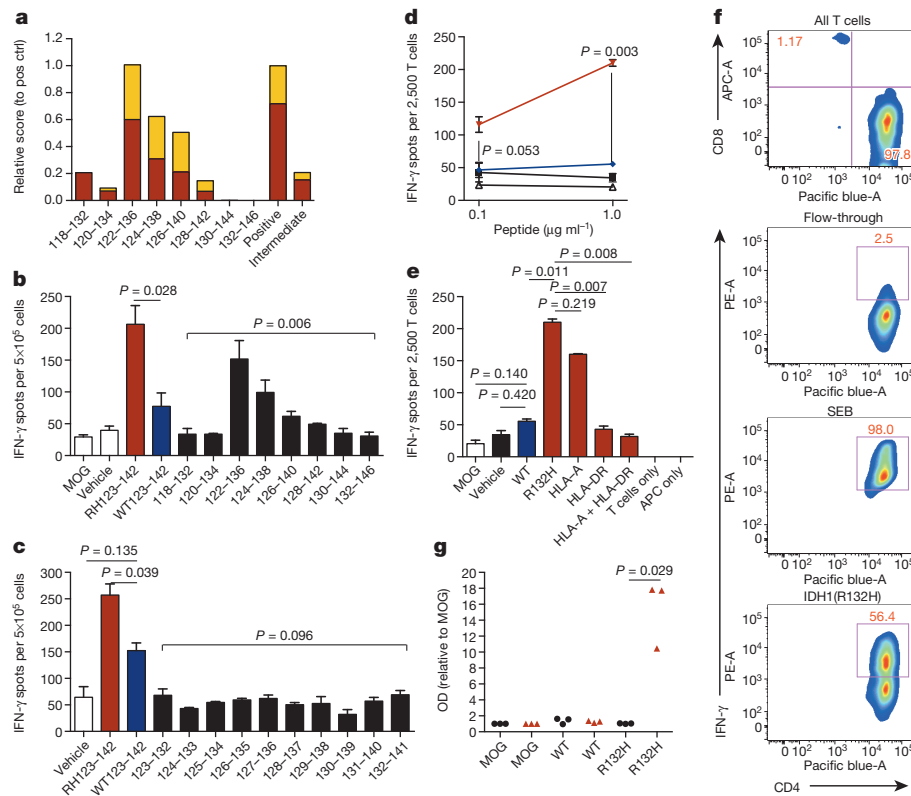
Data Fig. 1h, i). T-cell responses were restricted to long peptides, abrogated by an antibody against heterodimeric HLA-DRA (HLA-DR blocking antibody), and not induced by stimulation with 10-mer peptides, suggesting class II-restricted CD4<sup>+</sup> T-cell responses (Fig. 1b, c and Extended Data Fig. 1j) with a predominant T<sub>H</sub>1 phenotype as evidenced by an IDH1(R132H)-specific T-cell line (Fig. 1d–f, Extended Data Figs 2a–c and 3a–c, and Supplementary Note 1). CD4<sup>+</sup> T-cell responses are usually associated with an antibody response. Accordingly, in the serum of mice immunized with p123–142 (R132H) but not of control mice, mutation-specific anti-IDH1 (p123–142) antibodies were detectable (Fig. 1g and Extended Data Fig. 4a, b). Collectively, these data indicate that peptides covering the mutated region of human IDH1 induce a mutation-specific CD4<sup>+</sup> T-cell and antibody response in a human MHC context.

To assess clinical relevance, patients with gliomas and healthy controls (Extended Data Tables 1 and 2) were screened for spontaneous peripheral T-cell responses against IDH1(R132H). In 4 of 25 patients harbouring IDH1(R132H)-mutated gliomas but in none of 29 patients with an IDH1 wild-type glioma, IDH1(R132H)-specific IFN- $\gamma$ -producing T cells were detectable (Fig. 2a–c). As expected, the IDH1(R132H)-specific T-cell responses detectable in patients with IDH1(R132H)-mutated gliomas were not evoked by the 10-mer peptides (Fig. 2d and Extended Data Fig. 1a, b), were abrogated by an HLA-DR blocking antibody (Fig. 2e) and were predominantly CD4<sup>+</sup> (Fig. 2f). HLA typing of IDH1(R132H)-positive glioma patients revealed no clear restriction to a particular HLA type (DRB1\*1001;15, DRB1\*07;15, DRB1\*01;04 and DRB1\*03;13, respectively) (Supplementary Note 2 and Extended Data Table 1), supporting the notion that the class II epitope of IDH1(R132H) is promiscuous with respect to HLA class II type as suggested by the vaccination studies in HLA-DRB1\*0401 transgenic mice (DR4 mice) (Extended Data Fig. 2a–c). IDH1(R132H)-specific antibodies with predominance of IgG1 subclass were detected in 4 of 42 patients with IDH1(R132H) but not in 47 patients with IDH1 wild-type tumours and not in healthy donors (Fig. 2g, Extended Data Fig. 4c–e and Supplementary Note 2). Collectively, these data indicate that R132H-mutated IDH1 is naturally processed in glioma patients to present an immunodominant epitope in the p123–142 region on MHC class II molecules to CD4<sup>+</sup> T cells to induce a spontaneous mutation-specific T<sub>H</sub>1-polarized response and the production of mutation-specific antibodies detectable in patients with IDH1(R132H)-mutated but not IDH1 wild-type gliomas.

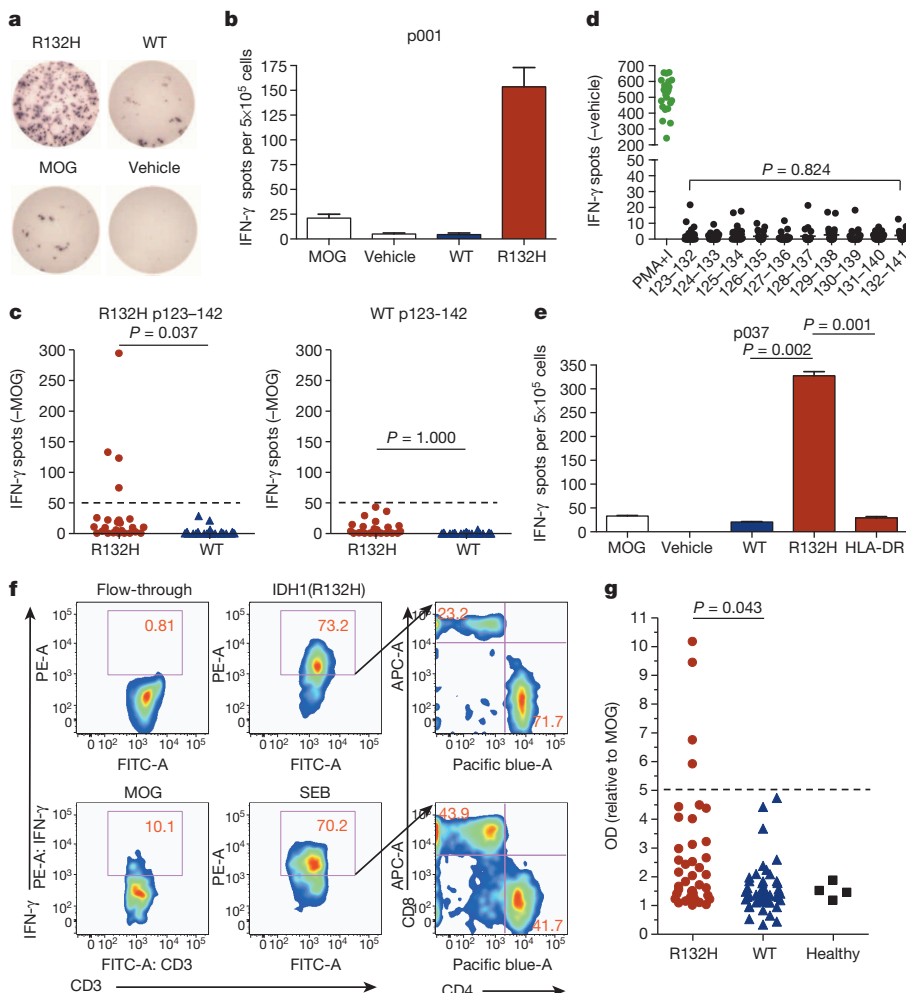
Next we analysed processing of the IDH1(R132H)-containing epitope and antitumour effects of the R132H-specific immune response. Whole-tumour cell vaccination of A2.DR1 mice with IDH1(R132H),

<sup>1</sup>Department of Neurooncology, University Hospital Heidelberg and National Center for Tumor Diseases, 69120 Heidelberg, Germany. <sup>2</sup>German Cancer Consortium (DKTK) Clinical Cooperation Unit Neuroimmunology and Brain Tumor Immunology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>3</sup>Department of Neuropathology, University Hospital Heidelberg and National Center for Tumor Diseases, 69120 Heidelberg, Germany. <sup>4</sup>German Cancer Consortium (DKTK) Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>5</sup>German Cancer Consortium (DKTK) Clinical Cooperation Unit Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>6</sup>Department of Translational Immunology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>7</sup>Department of Immunotherapy and -prevention Group, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>8</sup>Ribological GmbH, 55131 Mainz, Germany. <sup>9</sup>Translational Oncology, 55131 Mainz, Germany. <sup>10</sup>Department of Immunology, University of Tübingen, 72076 Tübingen, Germany. <sup>11</sup>Metabolic Centre Heidelberg, University Children's Hospital, 69120 Heidelberg, Germany. <sup>12</sup>Center for Molecular Neurobiology, University Medical Center, Hamburg-Eppendorf, 20251 Hamburg, Germany.

\*These authors contributed equally to this work.

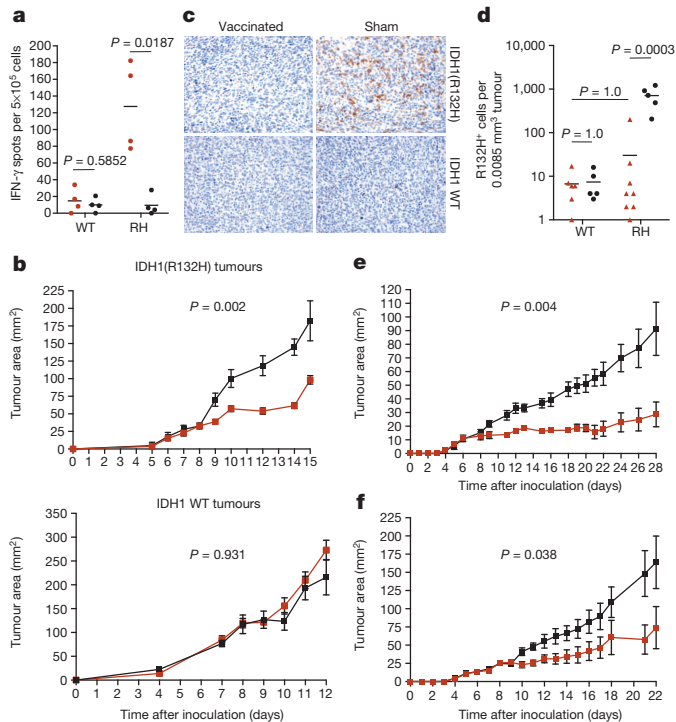


**Figure 1 | IDH1(R132H) peptide vaccination induces  $T_H1$  and humoral responses in A2.DR1 mice.** **a**, DR1-binding IDH1(R132H) 15-mer epitopes identified by Reveal Class II binding assay. Immediate binding (yellow) and 24-h stability (red) are shown. **b**, **c**, ELISpots of IFN- $\gamma$  splenocyte responses to 15-mers (**b**) and 10-mers (**c**) after vaccination with IDH1(R132H) in CFA (red, IDH1(R132H) (p123–142); blue, wild-type IDH1 (p123–142); black, IDH1(R132H) library; p123–142 (R132H versus wild type), Welch *t*-test; library, ANOVA;  $n = 3$  biological replicates). Error bars, mean  $\pm$  s.e.m. **d**, **e**, ELISpots (white, MOG; black, vehicle; red, IDH1(R132H); blue, wild-type IDH1; HLA-A, IDH1(R132H) plus HLA-A blocking antibody; HLA-DR, IDH1(R132H) plus HLA-DR blocking antibody; HLA-A + HLA-DR, IDH1(R132H) plus HLA-A and HLA-DR blocking antibodies; **d**, Welch *t*-test; R132H versus wild type;  $0.1 \mu\text{g ml}^{-1}$  wild type versus MOG,  $P = 0.41$ ;  $1.0 \mu\text{g ml}^{-1}$  wild type versus MOG,  $P = 0.20$ ;  $n = 3$ ; error bars, mean  $\pm$  s.e.m.; **e**, pairwise Welch *t*-test with Bonferroni correction;  $n = 2$ ; error bars, mean  $\pm$  s.e.m.). **f**, IFN- $\gamma$  secretion assay (flow through, IFN- $\gamma$ -negative fraction) of an IDH1(R132H)-specific T-cell line. Red numbers in **f** indicate populations in per cent. **n** values in **d** and **e** indicate number of technical replicates. **g**, Detection of IDH1(R132H)-specific IgG in serum after vaccination with IDH1(R132H) (red) and vehicle (black) using Montanide (Welch *t*-test;  $n = 3$  biological replicates). Scatter plot showing individual values. OD, optical density, absorbance.



**Figure 2 | Spontaneous  $T_H1$  and humoral IDH1(R132H)-specific responses are detected in IDH1(R132H)<sup>+</sup> glioma patients.** **a**, **b**, Representative ELISpots of IFN- $\gamma$  PBMC responses of patients p037 (**a**) and p001 (**b**) to IDH1 p123–142 (R132H and wild type (WT)). Error bars, mean  $\pm$  s.e.m.,  $n = 2$  technical replicates. **c**, **d**, ELISpots of IFN- $\gamma$  PBMC responses of patients with IDH1(R132H)-mutated gliomas (**c**, red,  $n = 25$ ; **d**,  $n = 23$ ) or wild-type IDH1 gliomas (**c**, blue,  $n = 29$ ) to IDH1 p123–142 (**c**, R132H and wild type) or R132H 10-mer peptides (**d**). Fisher's exact test (**c**). Welch ANOVA; PMA+I, phorbol myristate acetate + ionomycin, positive control (**d**). Scatter plots showing individual values (after subtraction of MOG-induced spots, negative values set to zero). Dashed line in **c** indicates cut-off for positivity. **e**, ELISpot of IFN- $\gamma$  PBMC responses of patient p037 to IDH1 p123–142 (red, R132H; blue, wild type) including HLA-DR block (Welch *t*-test;  $n = 3$  technical replicates). Error bars, mean  $\pm$  s.e.m. **f**, IFN- $\gamma$  secretion assay of PBMCs from patient p037 (flow through, IFN- $\gamma$ -negative fraction). Red numbers indicate populations in per cent. **g**, Detection of IDH1(R132H)-specific IgG in serum of patients with IDH1(R132H) ( $n = 42$ ), or wild-type IDH1 ( $n = 47$ ) gliomas, or healthy donors ( $n = 4$ ) (Fisher's exact test). Scatter plot showing individual values. Dashed line indicates cut-off for positivity set to 5.

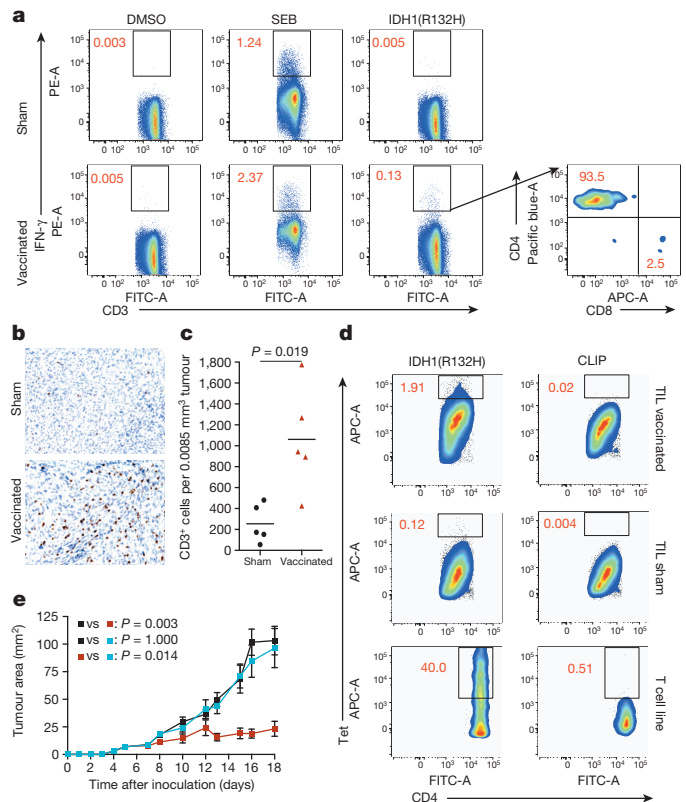




**Figure 3 | IDH1(R132H) peptide vaccination reduces IDH1(R132H)<sup>+</sup> tumour growth in A2.DR1 mice.** **a**, ELISpot of IFN- $\gamma$  splenocyte responses to IDH1 p123–142 (R132H (RH) or wild type (WT)) after whole-tumour cell (red, IDH1(R132H)<sup>+</sup>; black, wild-type IDH1<sup>+</sup>) vaccination (Welch *t*-test; *n* = 4). Scatter plot showing individual values (after subtraction of MOG-induced spots; negative values set to zero) and the mean. **b**, Growth of subcutaneous sarcomas (IDH1(R132H)<sup>+</sup> or wild-type IDH1<sup>+</sup>) after preventive vaccination with IDH1(R132H) (red) or vehicle (black) (Wilcoxon rank-sum test for median area under the curve (AUC); *n* = 8 IDH1(R132H), vaccination; *n* = 6 wild-type IDH1, vaccination; *n* = 5 IDH1(R132H) and wild type, sham). Error bars, mean  $\pm$  s.e.m. **c**, **d**, IDH1(R132H) expression in subcutaneous sarcomas (IDH1(R132H)<sup>+</sup> or wild-type IDH1<sup>+</sup>) after preventive vaccination with IDH1(R132H) (vaccination, red) or vehicle (sham, black). Representative images (c) and scatter plot showing individual values and the mean (d, Welch one-way ANOVA and pairwise Welch *t*-tests with Bonferroni correction after log<sub>2</sub> transformation; *n* as in b) are shown. **e**, **f**, Growth of pre-established subcutaneous tumours (e, IDH1(R132H)<sup>+</sup>; f, NY-ESO-1<sup>+</sup>) after vaccination with IDH1(R132H) (e) or NY-ESO-1 (f) (vaccination, red) or vehicle (black) (Wilcoxon rank-sum test for median AUC; *n* = 7). Error bars, mean  $\pm$  s.e.m. All *n* values indicate numbers of biological replicates.

but not IDH1 wild-type, A2.DR1 sarcomas (Extended Data Fig. 5a–e) induced a robust IDH1(R132H)-specific T-cell response, indicating that IDH1 is indeed endogenously processed to present the R132H-containing epitope on HLA-DRA\*0101 HLA-DRB1\*0101 (DR1) (Fig. 3a). An inducible antitumour immune response *in vivo* was demonstrated by vaccination of A2.DR1 mice with IDH1(R132H) (p123–142), which resulted in a growth suppression of subsequently transplanted IDH1(R132H), but not IDH1 wild-type, A2.DR1 sarcomas (Fig. 3b and Extended Data Fig. 5f, g). Importantly, IDH1(R132H)<sup>+</sup> tumours, which resisted vaccination, displayed greatly reduced IDH1(R132H) expression whereas sarcomas in the control group retained their IDH1(R132H) expression (Fig. 3c, d). The IDH1(R132H) peptide vaccine also suppressed the growth of pre-established IDH1(R132H)<sup>+</sup> sarcomas (Fig. 3e and Extended Data Fig. 5i, j) without any overt toxicity or impairment of IDH1 wild-type enzymatic activity (Extended Data Table 3 and Extended Data Fig. 5h). This therapeutic effect compared well with another HLA-DRB1\*0101-restricted peptide vaccine targeting a different, well-studied tumour-associated antigen: NY-ESO-1 (Fig. 3f, Extended Data Fig. 6a–d and Supplementary Note 3).

Although there is increasing evidence that antigen-specific cytotoxic CD4<sup>+</sup> T cells are in principle capable and sufficient to exert antitumour



**Figure 4 | Growth suppression of pre-established IDH1(R132H)<sup>+</sup> tumours by IDH1(R132H) peptide vaccination is dependent on T<sub>H</sub> cells.** **a**, IFN- $\gamma$  secretion assay of splenocytes pooled from 5 A2.DR1 mice vaccinated with IDH1(R132H) or vehicle (sham). **b**, **c**, CD3 expression in IDH1(R132H)<sup>+</sup> sarcomas after vaccination with IDH1(R132H) (red) or vehicle (sham, black). Representative images (b) and scatter plot showing individual values and the mean (c, Welch one-way ANOVA and pairwise Welch *t*-tests; *n* = 5). **d**, CD4<sup>+</sup> T cells in five pooled IDH1(R132H)<sup>+</sup> sarcomas (TIL) after vaccination with IDH1(R132H) or vehicle (sham) and IDH1(R132H)-specific T-cell line stained with IDH1(R132H) and control (CLIP) tetramers. **e**, Growth of subcutaneous IDH1(R132H)<sup>+</sup> sarcomas after depletion of CD4<sup>+</sup> T cells and vaccination with IDH1(R132H) (blue; depletion + vaccination; red, vaccination; black, sham; Wilcoxon rank-sum test for median AUC with Bonferroni correction; *n* = 6; *n* = 7 for vaccination). Error bars, mean  $\pm$  s.e.m. All *n* values indicate number of biological replicates.

immunity<sup>17,18</sup>, the relevance of a CD4<sup>+</sup> T-cell-mediated antitumour immune response in the absence of an MHC-class-I-restricted CD8<sup>+</sup> T-cell response is still controversial. Analyses of T cells recovered from IFN- $\gamma$  secretion assays after immunization of A2.DR1 mice with p123–142 (R132H) confirmed that CD4<sup>+</sup> but not CD8<sup>+</sup> T cells were antigen-specific (Fig. 4a and Extended Data Fig. 7a). In addition, IDH1(R132H)<sup>+</sup> A2.DR1 sarcomas were specifically infiltrated by IDH1(R132H) HLA-DRB1\*0101 tetramer-positive T cells (Fig. 4b–d). Finally, depletion of CD4<sup>+</sup> T cells abrogated the therapeutic effects of the IDH1(R132H) vaccine, indicating that the therapeutic efficacy is dependent on CD4<sup>+</sup> T cells (Fig. 4e and Extended Data Fig. 7b). Interestingly, the vaccine-mediated control of tumour growth was also abrogated when CD19<sup>+</sup> B cells were depleted, suggesting a contribution of B cells to the therapeutic efficacy of the IDH1(R132H) vaccine (Extended Data Fig. 7c–e and Supplementary Note 4).

In summary, we show that IDH1(R132H) represents a tumour-specific neoantigen recognized by CD4<sup>+</sup> IFN- $\gamma$ -producing T cells in patients with IDH1(R132H) tumours and that an IDH1(R132H) vaccine induces a specific antitumour immune response against IDH1(R132H)-mutated tumours in an MHC-humanized animal model. These data underline that mutant IDH may serve as a therapeutic target not only through drug-mediated inhibition of the neomorphic enzymatic

function<sup>19,20</sup> but also through T-cell-based targeting of the mutant epitope in a disease where the target is an early event in tumorigenesis and hence expressed in all tumour cells, which makes immunological escape unlikely to occur (Supplementary Note 5). Conceptually, patients with low-grade and anaplastic gliomas with a high prevalence of the IDH1(R132H) mutation represent a patient population that may particularly benefit from a tumour vaccine, because these tumours may remain stable or minimally growing for several years but will inevitably recur, often with a more malignant phenotype<sup>21</sup>, and because there is currently no maintenance therapy available preventing recurrence of this diffusely infiltrating disease in this relatively young and immunologically competent patient population<sup>22</sup>.

## METHODS SUMMARY

IDH1 peptide binding to HLA-A\*0201 and HLA-DRB1\*0101 was analysed by binding prediction algorithms NetMHC and SYFPEITHI, and *in vitro* T2 and class II REVEAL MHC-peptide binding assays. All animal procedures followed the institutional laboratory animal research guidelines and were approved by the governmental authorities. A2.DR1 and DR4 mice were vaccinated with IDH1(R132H) (p123–142) peptide in complete Freund's adjuvant (CFA) or Montanide-ISA51 with imiquimod and rmGM-CSF to assess the immunogenicity of IDH1(R132H) *in vivo* by cytokine analysis of splenocytes. IDH1(R132H)-specific CD4<sup>+</sup> T-cell line and clone were generated from splenocytes of vaccinated A2.DR1 mice by re-stimulations with autologous, irradiated and p132–142 (R132H)-loaded splenocytes and limiting dilution for in-depth analysis of IDH1(R132H) T-cell responses. To assess processing and presentation of IDH1(R132H) epitopes, a syngeneic sarcoma cell line was generated by chemical induction, transduced with IDH1(R132H) and wild type, irradiated and used for whole-tumour cell vaccination of A2.DR1 mice. Antitumour activity of induced IDH1(R132H) immune responses was shown by subcutaneous injection of syngeneic IDH1-expressing sarcoma cell line into previously or subsequently vaccinated A2.DR1 mice. Dependence on CD4<sup>+</sup> T-cell responses was shown by *in vivo* CD4 and CD19 depletion. IDH1 wild-type integrity was analysed by enzyme activity measurements in liver and brain. 2-HG production of cells and tumours was analysed enzymatically. Peripheral blood mononuclear cells (PBMCs) from patients with IDH1(R132H) and IDH1 wild-type gliomas were isolated to analyse spontaneous IDH1(R132H)-specific T-cell responses. IFN- $\gamma$  ELISpot, cytokine ELISA, intracellular flow cytometry, and IFN- $\gamma$ -secretion assays were used to assess cytokine profile and phenotype of specific T-cell responses after vaccination, in IDH1(R132H) T-cell line and clone, and of spontaneous T-cell responses in glioma patients. IDH1(R132H)-binding IgG was detected with peptide-coated ELISA in serum from vaccinated mice, patients and healthy donors. IDH1(R132H) expression was analysed by immunohistochemistry in human and mouse tumours and by western blot and immunofluorescence in mouse sarcoma cell lines.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 July 2013; accepted 17 April 2014.**

**Published online 25 June 2014.**

1. Balss, J. *et al.* Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta Neuropathol.* **116**, 597–602 (2008).
2. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
3. Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
4. Marcucci, G. *et al.* IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J. Clin. Oncol.* **28**, 2348–2355 (2010).
5. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
6. Amary, M. F. *et al.* IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *J. Pathol.* **224**, 334–343 (2011).

7. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
8. Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
9. Sasaki, M. *et al.* IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488**, 656–659 (2012).
10. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
11. Turcan, S. *et al.* IDH1 mutation is sufficient to establish the glioma hypermethylation phenotype. *Nature* **483**, 479–483 (2012).
12. Cairns, R. A. & Mak, T. W. Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities. *Cancer Discov.* **3**, 730–741 (2013).
13. Capper, D. *et al.* Characterization of R132H Mutation-specific IDH1 Antibody Binding in Brain Tumors. *Brain Pathol.* **20**, 245–254 (2010).
14. Watanabe, T., Nobusawa, S., Kleihues, P. & Ohgaki, H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am. J. Pathol.* **174**, 1149–1153 (2009).
15. Capper, D., Zentgraf, H., Balss, J., Hartmann, C. & von Deimling, A. Monoclonal antibody specific for IDH1 R132H mutation. *Acta Neuropathol.* **118**, 599–601 (2009).
16. Pajot, A. *et al.* A mouse model of human adaptive immune functions: HLA-A2.1-/HLA-DR1-transgenic H-2 class I/class II-knockout mice. *Eur. J. Immunol.* **34**, 3060–3069 (2004).
17. Hunder, N. N. *et al.* Treatment of metastatic melanoma with autologous CD4<sup>+</sup> T cells against NY-ESO-1. *N. Engl. J. Med.* **358**, 2698–2703 (2008).
18. Quezada, S. A. *et al.* Tumor-reactive CD4<sup>+</sup> T cells develop cytotoxic activity and eradicate large established melanoma after transfer into lymphopenic hosts. *J. Exp. Med.* **207**, 637–650 (2010).
19. Wang, F. *et al.* Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626 (2013).
20. Rohle, D. *et al.* An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* **340**, 626–630 (2013).
21. Wick, W. *et al.* NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide. *J. Clin. Oncol.* **27**, 5874–5880 (2009).
22. Jansen, M., Yip, S. & Louis, D. N. Molecular pathology in adult gliomas: diagnostic, prognostic, and predictive markers. *Lancet Neurol.* **9**, 717–726 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are indebted to the patients and their relatives who agreed to participate in this study. We thank A. Gardyan, T. Lanz and W. Osen for technical advice, T. Lanz and A. Hertenstein for providing patient blood samples, S. Hundt for cloning, A. Habel and D. Krunic for technical support, W. Nicklas for pathological analysis, and J. Jung for graphics design. We acknowledge the support by the DKFZ Light Microscopy and Genomics and Proteomics Facilities. HLA-DRB1\*01:01 MHC class II tetramer bound to IDH1(R132H) p123–142 was provided by NIH tetramer core facility. A2.DR1 mice were provided by Institut Pasteur. This work was supported by the Interdisciplinary Research Program of the National Center for Tumor Diseases Heidelberg (IFP III/2) to M.P. and A.V.D., the Wilhelm Sander Foundation (2012.118.1) to M.P. and A.V.D., the Helmholtz Foundation (VH-NG-306) and the Andreas Zimprich Foundation to M.P. and the German Research Foundation (SFB938 TPK) to M.P. and W.W. T.S. and M.K. were supported by the Helmholtz International Graduate School, and L.B. was supported by the Heinrich F. C. Behr Foundation and the Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology MD/PhD program, University Heidelberg. F.S. was supported by a postdoctoral fellowship of the University Hospital Heidelberg.

**Author Contributions** T.S. and L.B. designed and performed experiments, analysed data and wrote the paper. F.S. and A.V.D. provided glioma tissue and determined IDH1 status. S.P. cloned IDH1 constructs. J.B. and J.O. performed 2-HG measurements. J.Q. and P.B. generated the A2.DR1 sarcoma cell line. K.R. and M.O. established stable overexpressions. B.W. performed statistical analysis. I.O. and M.K. performed animal experiments. O.M. and M.Os. provided patient blood samples. A.K.G. and A.B.R. performed epitope prediction and T2 binding assays. J.D., N.T., I.V., S.S. and U.S. analysed patient samples. S.B.E. provided DR4 mice. M.A.F. and W.W. were involved in study design and data interpretation. M.P. conceptualized the study, designed experiments, interpreted data and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P. (m.platten@dkfz.de).

# Dynamic pathways of $-1$ translational frameshifting

Jin Chen<sup>1,2</sup>, Alexey Petrov<sup>2</sup>, Magnus Johansson<sup>2</sup>, Albert Tsai<sup>1,2</sup>, Seán E. O'Leary<sup>2</sup> & Joseph D. Puglisi<sup>2</sup>

Spontaneous changes in the reading frame of translation are rare (frequency of  $10^{-3}$  to  $10^{-4}$  per codon)<sup>1</sup>, but can be induced by specific features in the messenger RNA (mRNA). In the presence of mRNA secondary structures, a heptanucleotide 'slippery sequence' usually defined by the motif X XXY YYZ, and (in some prokaryotic cases) mRNA sequences that base pair with the 3' end of the 16S ribosomal rRNA (internal Shine–Dalgarno sequences), there is an increased probability that a specific programmed change of frame occurs, wherein the ribosome shifts one nucleotide backwards into an overlapping reading frame ( $-1$  frame) and continues by translating a new sequence of amino acids<sup>2,3</sup>. Despite extensive biochemical and genetic studies, there is no clear mechanistic description for frameshifting. Here we apply single-molecule fluorescence to track the compositional and conformational dynamics of individual ribosomes at each codon during translation of a frameshift-inducing mRNA from the *dnaX* gene in *Escherichia coli*. Ribosomes that frameshift into the  $-1$  frame are characterized by a tenfold longer pause in elongation compared to non-frameshifted ribosomes, which translate through unperturbed. During the pause, interactions of the ribosome with the mRNA stimulatory elements uncouple EF-G catalysed translocation from normal ribosomal subunit reverse-rotation, leaving the ribosome in a non-canonical intersubunit rotated state with an exposed codon in the aminoacyl-tRNA site (A site). tRNA<sup>Lys</sup> sampling and accommodation to the empty A site and EF-G action either leads to the slippage of the tRNAs into the  $-1$  frame or maintains the ribosome into the 0 frame. Our results provide a general mechanistic and conformational framework for  $-1$  frameshifting, highlighting multiple kinetic branchpoints during elongation.

Despite detailed biochemical and genetic studies, the mechanism of  $-1$  programmed ribosomal frameshifting (PRF) remains poorly understood, with at least three groups of models attempting to explain frameshifting<sup>2,4–11</sup>. The 3'-hairpin<sup>12</sup> and 5'-internal-Shine–Dalgarno sequence<sup>13</sup> (in some prokaryotic cases) pause the ribosome over the slippery sequence, which is necessary but not sufficient to drive efficient  $-1$  PRF. How these structural elements induce the pause and operate together to manipulate the ribosomal reading frame are not known. During a  $-1$  frameshift, the anticodons of the tRNAs must detach from the mRNA and re-associate in the  $-1$  frame. However, this slippage may occur at distinct points during the elongation cycle: (1) during accommodation of the A-site tRNA<sup>4</sup>, (2) subsequent to accommodation, but before peptidyl transfer<sup>8</sup>, (3) during EF-G catalysed translocation<sup>9,10</sup>, or (4) after translocation but before the next round of elongation<sup>11</sup>. As the exact timing of frameshifting is unknown, the precise position of the ribosome over the slippery sequence during the slippage is unclear<sup>8,11</sup>. Finally, as  $-1$  PRF has been shown to occur at approximately 1% to 80% efficiency depending on the sequence<sup>10</sup>, what determines whether one particular ribosome will frameshift or not remains elusive.

The dynamic and stochastic nature of frameshifting requires direct observation of single ribosomes translating multiple codons of an mRNA. We harness here single-molecule fluorescence and zero-mode waveguides (ZMWs) instrumentation<sup>14</sup> to track ribosome progression on mRNAs and observe  $-1$  translational frameshifting in real-time<sup>15</sup>. Conformational changes underlying elongation, involving rotational movements of the small (30S) ribosomal subunit body with respect to the large (50S)

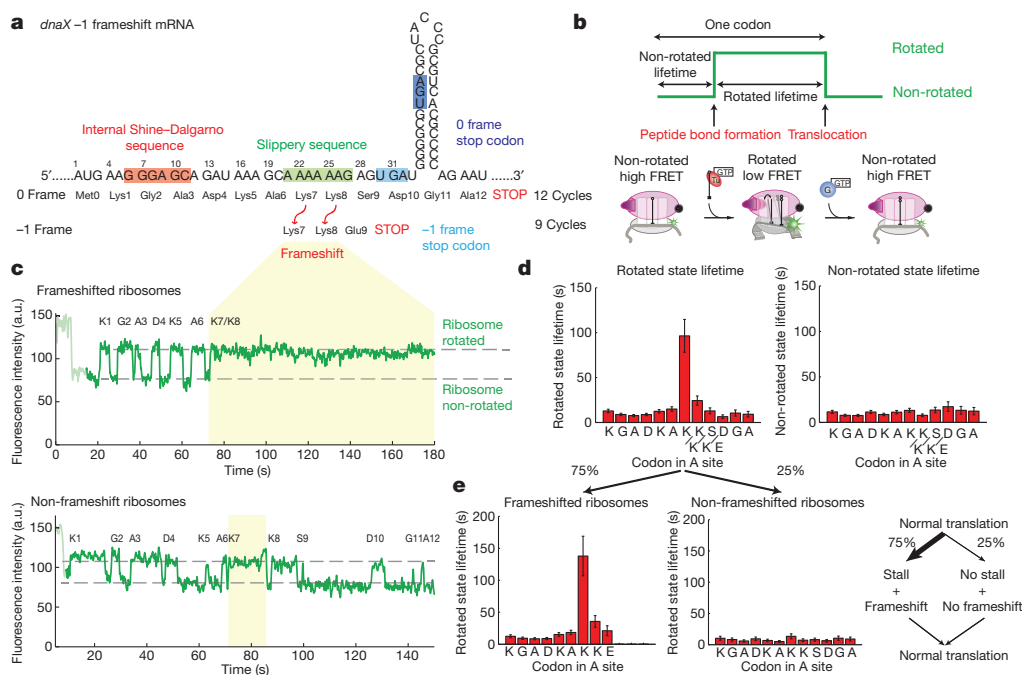
ribosomal subunit, were monitored during translation by site-specifically labelling the 30S with Cy3B and 50S with BHQ-2 (a non-fluorescent quencher), allowing for Förster resonance energy transfer (FRET) between the two dyes<sup>16,17</sup>. During normal translation elongation, aminoacyl-tRNA–EF-Tu–GTP ternary complex accommodation to the A site followed by peptide bond formation drives the non-rotated to rotated state transition (low to high Cy3B intensity, or high to low FRET), whereas EF-G catalysed translocation drives the rotated to non-rotated transition (high to low Cy3B intensity)<sup>16</sup>. Thus, one round of high-low-high FRET (low-high-low Cy3B intensity) corresponds to a single ribosome translating one codon (see Extended Data Fig. 1)<sup>16,18</sup>. Arrival and departure of the dye-labelled ligands such as Cy5–tRNAs and Cy5–EF-G can be simultaneously observed as a sequence of fluorescent pulses<sup>16,17</sup>. We applied this approach to the  $-1$  frameshift sequence from the *dnaX* gene in the *E. coli*, which contains an internal Shine–Dalgarno sequence and a slippery -A AAA AAG- sequence followed by a RNA hairpin<sup>3,12,13</sup>.

We observed  $-1$  frameshifting directly on a *dnaX* frameshift sequence, designed such that ribosomes that frameshift will translate 9 codons and stop at a stop codon in the  $-1$  frame, whereas ribosomes that do not frameshift will translate 12 codons until a stop codon in the 0 frame (Fig. 1a). By delivering total tRNA (tRNA<sub>tot</sub>) ternary complex, EF-G, and BHQ-50S to immobilized Cy3B-30S preinitiation complexes (30S subunit–mRNA–initiator tRNA), we observe ribosomes that translate either the full 12 codons or only 9 codons, as measured by the number of intersubunit FRET cycles (see Extended Data Fig. 2). By determining the fraction of ribosomes that translate  $> 9$  codons, or translate up to 9 codons, we obtain an estimate of the frameshifting percentage (75%), consistent with previously observed frameshifting efficiency<sup>13</sup> (confirmed independently as shown in Extended Data Fig. 2b, c). The Shine–Dalgarno sequence and hairpin act as barriers to translocation, so mutations of the potential Shine–Dalgarno sequence and removal of the hairpin all decrease frameshifting efficiency as expected (Extended Data Figs 3 and 4).

Elongation of the *dnaX* mRNA is drastically and abruptly perturbed at the seventh FRET cycle (codon Lys7). Analysis of rates at each codon revealed a tenfold increase in the rotated state (waiting for EF-G and translocation) lifetime ( $96 \pm 18$  s vs  $\sim 5$ – $10$  s for the other codons) at Lys7, corresponding to tRNA<sup>Ala</sup>(GCA<sub>21</sub>)-codon pair in the ribosomal peptidyl-tRNA site (P site) and the newly incorporated tRNA<sup>Lys</sup>(AAA<sub>24</sub>) codon pair in the A site (GCA<sub>21</sub> AAA<sub>24</sub>), poised for translocation; non-rotated state lifetimes (waiting for ternary complex and peptide bond formation) remain constant at each codon (Fig. 1b–d). Furthermore by partitioning frameshifted vs non-frameshifted ribosomes, an increased rotated-state lifetime at codon Lys7 is observed only for frameshifted ribosomes ( $138 \pm 31$  s); non-frameshifted ribosomes translate through the frameshift site seemingly unaffected ( $13 \pm 4$  s) (Fig. 1e, confirmed independently in Extended Data Fig. 2d, and repeated with varying factor concentrations in Extended Data Fig. 5). Disruption of the slippery sequence by changing A<sub>21</sub> AAA<sub>24</sub> AAG<sub>27</sub> to G<sub>21</sub> AAG<sub>24</sub> AAG<sub>27</sub> (A21G–A24G mutant) caused an expected decrease in frameshifting efficiency to 12% (background level in our experiments is  $\sim 3$ – $10\%$ ), while drastically decreasing the lifetime at codon Lys7 ( $25 \pm 5$  s instead of  $96 \pm 18$  s) (Extended Data Fig. 6). Thus, the long lifetime at codon Lys7 is a hallmark of frameshifting and requires the slippery-site sequence. Partitioning between frameshifted and non-frameshifted ribosomes was assumed to

<sup>1</sup>Department of Applied Physics, Stanford University, Stanford, California 94305-4090, USA. <sup>2</sup>Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305-5126, USA.





**Figure 1 | Frameshifting is characterized by a long rotated-state pause.**

**a**, Schematic of the mRNA used in this study, modified from the *dnaX* gene. **b**, Schematic of the Cy3B/BHQ ribosome FRET signal, with each low-high-low Cy3B intensity cycle representing a ribosome elongating one codon. **c**, Sample traces of Cy3B (green) fluorescent intensity for frameshifted and non-frameshifted ribosomes translating with 80 nM EF-G and 1  $\mu$ M tRNA<sub>tot</sub> ternary complex. Codon Lys7 of the frameshift site is shaded yellow.

occur during the pause induced by frameshift signal. Instead, we demonstrate that the initial branch point occurs before the pause, but all frameshifted ribosomes exhibit a pause.

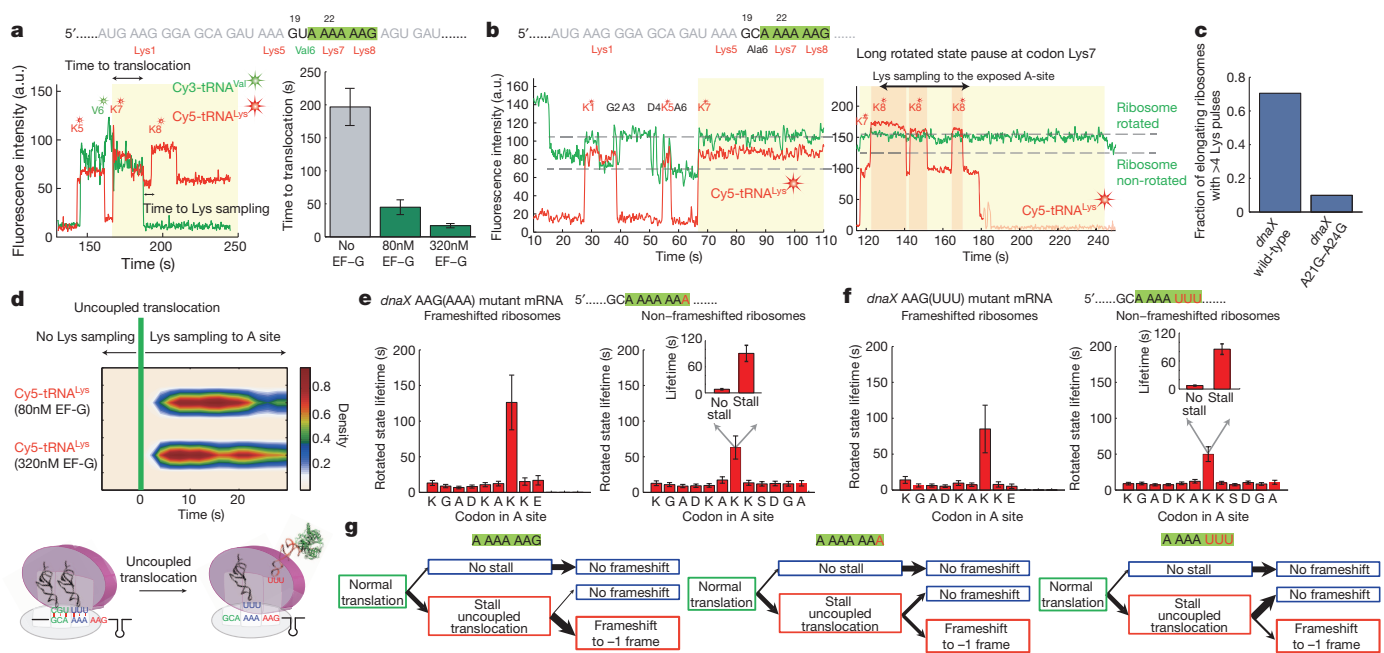
We next determined what is occurring during the pause that is characteristic of frameshifting. Normally translocation is coupled to ribosome reverse-rotation with deacylated tRNA in the ribosomal exit site (E site) departing rapidly after the ribosome reverse-rotates<sup>16</sup>. Using Cy3-labelled tRNA<sup>Val</sup>, we observed E-site tRNA departure directly at the frameshift site on a GCA<sub>21</sub> (Ala) to GUA<sub>21</sub> (Val) mRNA mutant, without affecting the frameshifting behaviour (Extended Data Fig. 7). We measured the departure of Cy3-tRNA<sup>Val</sup> relative to the Cy5-tRNA<sup>Lys</sup> arrival to the AAA<sub>24</sub> (Lys7) codon in the A site, which defines the start of the long rotated-state pause, correlated to peptide bond formation and transition to the rotated state: departure of deacylated Cy3-tRNA<sup>Val</sup> relative to the arrival of Cy5-tRNA<sup>Lys</sup> at codon Lys7 estimates when and if translocation occurs during the pause. During translation of the *dnaX* mRNA, Cy3-tRNA<sup>Val</sup> departs on average  $45 \pm 11$  s after the arrival of Cy5-tRNA<sup>Lys</sup> to the Lys7 codon (within a photobleaching time of  $196.7 \pm 28.1$  s). This time decreases with increasing concentration of EF-G, confirming that tRNA departure is linked to translocation (Fig. 2a). However, as the Cy3-tRNA<sup>Val</sup> residence time is much shorter than the rotated state lifetime (138 s), translocation occurs within the rotated state pause and precedes eventual reverse rotation. Thus translocation at Lys7 during frameshifting is uncoupled from reverse rotation of the ribosomal subunits and moves the ribosome A and P sites over the slippery sequence (AAA<sub>24</sub> AAG<sub>27</sub>). Translocation in this case is still inhibited through the interactions with the hairpin and internal Shine-Dalgarno sequence, with time to translocation longer than normal translation.

Uncoupling of tRNA-mRNA translocation from reverse-rotation and E-site tRNA departure creates a non-canonical intermediate in translation: the ribosome has a peptidyl-tRNA in the P site, but remains in a rotated intersubunit conformation. To delve into the nature of this intermediate frameshifting state and whether the A site is available for tRNA binding, we correlated Cy5-tRNA<sup>Lys</sup> binding and departure events with

a.u., arbitrary units. **d**, The mean rotated-state lifetime and non-rotated state lifetime for each FRET cycle. The non-rotated state lifetime is constant. There is a tenfold increase in rotated state lifetime at codon Lys7 (note that this lifetime consists of codon Lys7 and Lys8 due to uncoupled translocation, see Fig. 2).  $n = 256$ . Error bars, s.e. **e**, By parsing the rotated state lifetimes in **c** into ribosomes that frameshift (75%) and ribosomes that do not frameshift (25%), we see only the rotated state pause for frameshifted ribosomes.

the ribosome FRET signal (Fig. 2b). Although the *dnaX* mRNA sequence consists of 4 Lys codons, 71% of elongating ribosomes exhibit  $> 4$  Cy5-tRNA<sup>Lys</sup> pulses (Fig. 2c) (equal to the frameshifting percentage). The first three tRNA<sup>Lys</sup> pulses (Lys1, Lys5 and Lys7) show arrival rates and lifetimes consistent with elongation dynamics from intersubunit FRET data (Extended Data Fig. 8a, b); the third Lys7 pulse corresponds to the ribosome decoding AAA<sub>24</sub> at the slippery site (lifetime of 119.4 s, consistent with the rotated state lifetime at Lys7). The existence of the fourth and subsequent tRNA pulses directly indicate that translocation has occurred during the long rotated state and the A site is now available for aminoacyl-tRNA binding. After uncoupled translocation of the tRNA<sup>Lys</sup> to the P site, which would expose the fourth Lys codon (Lys8), tRNA<sup>Lys</sup> samples the A-site codon multiple times (on average 2.3 times), resulting in a buildup of Cy5 intensity (from two Cy5-tRNA<sup>Lys</sup> bound to the ribosome, Fig. 2b, c) even though the rotated state is not the natural substrate for tRNA binding to the A site. Mutation of the slippery sequence (A21G-A24G) greatly suppresses additional sampling by Cy5-tRNA<sup>Lys</sup> (only 9.9% of elongating ribosomes exhibit  $> 4$  pulses), indicating that multiple sampling events on Lys8 are characteristic of frameshifting and the long pause. Post-synchronization of the arrival of the fourth sampling tRNA<sup>Lys</sup> to the time of uncoupled translocation shows that translocation gates the arrival of the sampling tRNA<sup>Lys</sup>, confirming that tRNA<sup>Lys</sup> is indeed sampling the A-site codon exposed by translocation (Fig. 2d and Extended Data Fig. 8). Delivery of tRNA-EF-Tu-GDPNP (a non-hydrolysable analogue of GTP) instead of GTP decreases the tRNA pulse lifetimes from  $38 \pm 2$  s to  $2.1 \pm 0.1$  s, demonstrating that GTP hydrolysis by EF-Tu and subsequent accommodation of the tRNA into the ribosomal A site occur for these long-lived sampling pulses.

We propose that uncoupled +3 translocation creates weakened codon-anticodon-ribosome interactions, and that tRNA<sup>Lys</sup>-sampling and accommodation at the AAG<sub>27</sub> codon presented by the non-canonical rotated ribosome drives the ribosome into the -1 frame and helps re-establish codon-anticodon interactions (see Extended Data Fig. 9)<sup>4,8,10</sup>. Consistent with this model, AAA<sub>26</sub> (-1 frame) and AAG<sub>27</sub> (0 frame) both



**Figure 2 | tRNA samples the rotated state after uncoupled translocation and defines the reading frame.** **a**, Sample trace and time to translocation at the frameshift site, which is estimated by Cy3-tRNA<sup>Val</sup> (green) departure from the E site during frameshifting relative to the arrival of Cy5-tRNA<sup>Lys</sup> (red) at codon Lys7 (shaded in yellow) on a GCA<sub>21</sub> (Ala) to GUA<sub>21</sub> (Val) mRNA mutant. As time to translocation is shorter than the rotated state pause, translocation from Lys7 to Lys8 occurs during the pause and is uncoupled from reverse-rotation. From left to right,  $n = 337$ ,  $n = 449$  and  $n = 455$ . Error bars, s.e. **b**, Sample trace of correlation of the Cy3B/BHQ ribosome FRET signal (green) with Cy5-tRNA<sup>Lys</sup> (red), confirming the long pause at codon Lys7 (shaded yellow). Upon reaching codon Lys7, additional tRNA<sup>Lys</sup> pulses sample codon Lys8 in the A site (shaded red), which results in a buildup of Cy5 intensity from the two Cy5-tRNA<sup>Lys</sup> in the A and P sites of the ribosome.

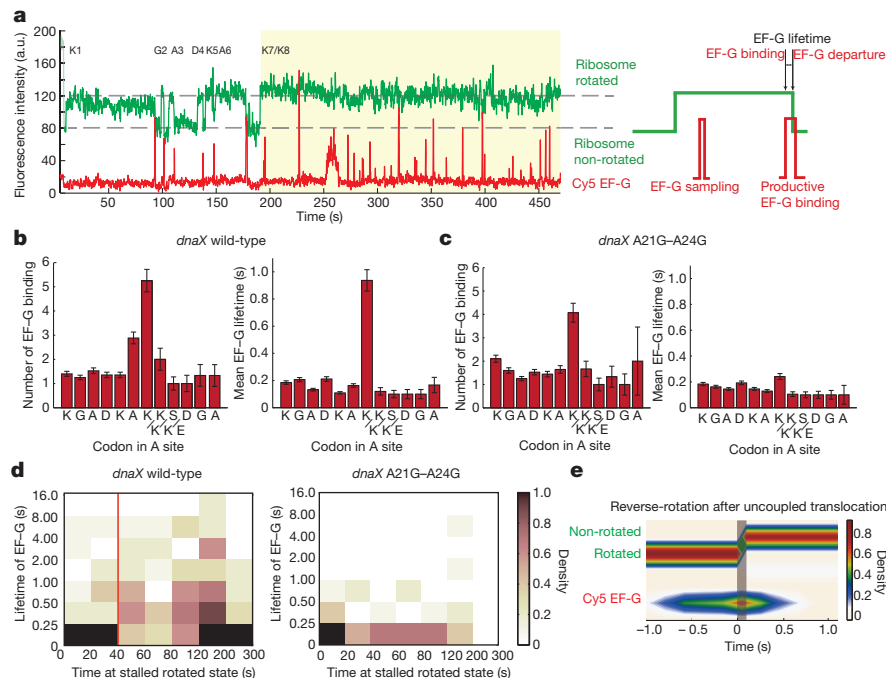
encode Lys, but an AAG codon interacts less stably than an AAA codon with the UUU anticodon of Lys-tRNA<sup>Lys</sup> because of tRNA modifications and wobble pairing at the third position<sup>19–21</sup>, favouring simultaneous slippage of the two codon-anticodon interactions in the  $-1$  direction. Mutation of the AAG<sub>27</sub> codon to AAA<sub>27</sub> (AAG(AAA) mutant) removes this preference, correspondingly decreasing the frameshift percentage to 49%<sup>19</sup>. For the frameshifted ribosomes on the AAG(AAA) mutant, the characteristic long rotated-state stall (mean lifetime  $126 \pm 38$  s) is still observed, with tRNA<sup>Lys</sup> sampling driving to the  $-1$  frame, as for the wild-type mRNA. However, there are two populations for the non-frameshifted ribosomes: one without the long pause (mean lifetime  $8.3 \pm 1.4$  s) in which ribosomes translate unaffected through the frameshift site (as before), and a second new subpopulation with a long pause at codon Lys7 (mean lifetime  $90.8 \pm 18.6$  s) that remains in the 0 frame, with tRNA<sup>Lys</sup> sampling to the 0 frame (Fig. 2e). We also observe similar behaviour when changing AAG<sub>27</sub> to UUU<sub>27</sub> (AAG(UUU)), eliminating possible slippage between the A site and P site tRNA-codon interactions, which decreases the frameshifting percentage to 20% (Fig. 2f). These results suggest that the long-lived rotated state, with P-site tRNA and an empty A-site codon is not sufficient for frameshifting; there is an additional need for slippage between the A- and P-site tRNA codon-anticodon interactions. The final reading frame is established after uncoupled translocation, and that long-lived sampling of the A site by tRNA<sup>Lys</sup> serves to define reading frame before the non-canonical state can be resolved. Thus, there is a second branch point for reading frame determination during  $-1$  frameshifting, which involves tRNA sampling in the A site (Fig. 2g).

To examine how the long-lived rotated state is resolved to continue translation once reading frame is established, we delivered Cy5-EF-G and correlated the EF-G pulses with the ribosome FRET signal (Fig. 3a). During the long rotated state, EF-G samples the ribosome multiple times<sup>16</sup> (on

**c**, Fraction of elongating ribosomes exhibiting  $> 4$  Lys pulses (additional sampling pulses) for the dnaX wild-type mRNA and the A21G-A24G mutant. From left to right,  $n = 179$  and  $n = 147$ . **d**, Two-dimensional density plot of Cy5-tRNA<sup>Lys</sup> sampling to codon Lys8 in the A site post-synchronized to the time of uncoupled translocation (indicated by the green line). The sampling of tRNA<sup>Lys</sup> at Lys8 only begins after translocation. **e**, Rotated and non-rotated state lifetimes for the slippery sequence mutant (A<sub>25</sub>A<sub>26</sub>G<sub>27</sub> codon to AAA). There is now an extra subpopulation of ribosomes with a long rotated-state pause that does not lead to frameshifting.  $n = 310$ , error bars, s.e. **f**, Mutation of the last A<sub>25</sub>A<sub>26</sub>G<sub>27</sub> codon to UUU (Phe). Similar to **e**, there are two subpopulations within the non-frameshifted ribosomes.  $n = 353$ , error bars, s.e. **g**, Pathways of frameshifting for the various slippery sequence mutants, indicating how tRNA sampling defines the final reading frame.

average  $> 5$  times per codon vs 1.3 per codon for non-frameshifting codons). Multiple long EF-G pulses (mean lifetime of  $\sim 1$  s vs  $\sim 100$  ms for non-frameshifting codons) are observed at codon 7. These are not observed if the slippery site is mutated; instead multiple short EF-G sampling events occur to allow translocation (Fig. 3b, c), consistent with an increased energy barrier imposed by the hairpin and Shine-Dalgarno interactions. For the wild-type mRNA, the long lifetime EF-G pulses occur most frequently during and after the uncoupled translocation to codon Lys8 (after  $\sim 40$  s) (Fig. 3d), echoing a cryo-EM structure of a eukaryotic ribosome over a  $-1$  frameshifting signal in the rotated state showing that the eukaryotic equivalent of EF-G, eEF2, was trapped on the ribosome<sup>9,22</sup>. The long-paused state is finally resolved by EF-G-GTP, as revealed by post-synchronization of ribosome reverse-rotation correlated with occupancy of Cy5-EF-G (Fig. 3e). Thus, EF-G action resolves the frameshift state and resumes translation, and may have a role during tRNA<sup>Lys</sup> sampling to promote the  $-1$  slippage.

Here we directly tracked translation in real time to follow the dynamics of frameshifting, developing a mechanistic model that embraces prior biochemical and structural studies<sup>9,23</sup>. We propose that the stochastic interaction of the ribosome with the hairpin helix in an open or closed state, and/or formation of the Shine-Dalgarno and anti-Shine-Dalgarno pairing interaction represent the shunt to either pausing in the rotated state (which leads to uncoupled translocation) or normal translation<sup>24</sup>. The long-lived, rotated ribosomal state contains a peptidyl-tRNA<sup>Lys</sup> in the P site and AAG<sub>27</sub> codon in the A site creating a non-canonical intermediate in translation, which is required for frameshifting and probably involves weakened tRNA-mRNA-ribosome contacts<sup>25,26</sup> and may involve hyper-rotation as recently proposed<sup>27</sup>. This state frustrates the normal action of both tRNA-EF-Tu-GTP and EF-G-GTP, accounting for the long pause. The repeated sampling of tRNAs to this state during the long pause allows

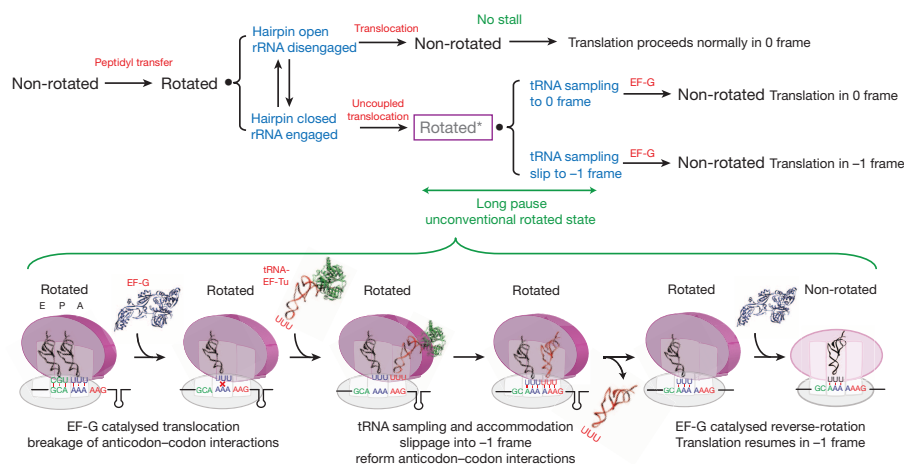


**Figure 3 | EF-G samples and resolves the uncoupled rotated state after frameshifting.** **a**, Sample trace and schematic of the correlation of Cy3B/BHQ ribosome FRET signal (green) with Cy5-EF-G binding (red). Cy5-EF-G pulses are correlated with ribosome reverse-rotation at each codon. At the rotated state pause (shaded in yellow), multiple EF-G sampling events with long dwell times can be observed. **b**, EF-G sampling and EF-G lifetimes for each codon for the wild-type frameshift mRNA. There is an increased number of sampling events as well as increased mean EF-G lifetime during the seventh FRET cycle (codon Lys7 and Lys8 due to uncoupled translocation).  $n = 122$ ; error bars, s.e.

**c**, For the A21G–A24G mutant mRNA, there is a slight increase in number of sampling events at Lys7 codon due to the increase in energy barrier from the hairpin and internal Shine–Dalgarno interaction, but the long EF-G lifetime disappears.  $n = 157$ , error bars, s.e. **d**, Two-dimensional histogram plotting time at the stalled rotated state vs lifetime of EF-G. Longer EF-G lifetimes only appear after uncoupled translocation, as roughly indicated by the red line. For the A21G–A24G mutant, no long EF-G lifetimes are observed.  $n = 122$  (left),  $n = 157$  (right). **e**, Post-synchronization plot correlating ribosome reverse-rotation after the rotated-state pause with EF-G.  $n = 436$ .

the binding energy of the codon–anticodon pairing to be used to allow slippage, whereas both P- and A-site tRNAs are on the ribosome and redefine the translational frame; EF-G may facilitate slippage. As recently proposed, the tRNA hybrid states may be destabilized, favouring a classical-like conformation of the two tRNAs that promotes  $-1$  slippage<sup>15</sup>. Peptidyl transfer is probably inefficient, as the rotated ribosome probably does not position the two tRNAs correctly for peptidyl transfer to occur

efficiently. EF-G eventually resolves the state and continues translation. The competition between slow peptide-bond formation and slow translocation explains heterogeneous protein products in prior frameshifting studies<sup>28</sup> (Extended Data Fig. 10). Thus, frameshifting involves both EF-G and tRNA, and occurs at an unconventional point during elongation after translocation of the ribosome on to the slippery sequence (uncoupled with reverse-rotation) but before peptidyl transfer<sup>11</sup> (Fig. 4), though



**Figure 4 | Branchpoint of pathways and mechanism of *dnaX*  $-1$  frameshifting.** The first branchpoint during frameshifting is probably due to the stochastic interaction of the ribosome with the hairpin in an open or closed state, and/or formation of the Shine–Dalgarno and anti-Shine–Dalgarno pairing, that represent the shunt to either pausing or normal translation. Translocation of the paused ribosomes under the slippery sequence with the tension caused by the hairpin and Shine–Dalgarno leads to  $+3$  translocation,

but uncoupled from reverse ribosomal rotation, creating a non-canonical intermediate in translation (denoted Rotated\*). The uncoupled translocation exposes the A site, to which tRNA<sup>Lys</sup> and EF-G sample. tRNA<sup>Lys</sup> sampling and accommodation to the AAG codon and EF-G action stimulates the ribosome to slip into the  $-1$  frame. Finally, EF-G catalyses the final reverse-rotation, after which the ribosome resumes normal translation. Depending on the system and condition, other pathways probably exist.



multiple pathways probably exist<sup>29</sup>. The interplay of mRNA sequence and structure with ribosomal dynamics leads to branchpoints during elongation, creating non-canonical paused states that allow unusual events in elongation. Such states may be a central feature of translational control.

## METHODS SUMMARY

All labelled ribosomes, factors, and tRNAs were prepared and purified as described<sup>16</sup>. Unless noted otherwise, all experiments were performed under buffer conditions described in Methods. Data collection from ZMW chips was conducted using instrumentation and techniques described previously<sup>14,30</sup>. Fluorescence traces were recorded at 10 frames per second for 8 min, with delivery of ligands to start the experiment at  $t = 10$  s. Statistical analysis on those traces was also conducted as described before<sup>16,30</sup>, using a custom software written in MATLAB (MathWorks). All error bars presented on figures are standard errors (s.e.).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 February; accepted 1 May 2014.**

**Published online 11 June; corrected online 20 August 2014 (see full-text HTML version for details).**

- Jenner, L. B., Demeshkina, N., Yusupova, G. & Yusupov, M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nature Struct. Mol. Biol.* **17**, 555–560 (2010).
- Tinoco, I., Jr, Kim, H. K. & Yan, S. Frameshifting dynamics. *Biopolymers* **99**, 1147–1166 (2013).
- Tsuchihashi, Z. & Kornberg, A. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl Acad. Sci. USA* **87**, 2516–2520 (1990).
- Plant, E. P. *et al.* The 9-Å solution: how mRNA pseudoknots promote efficient programmed –1 ribosomal frameshifting. *RNA* **9**, 168–174 (2003).
- Baranov, P. V., Gesteland, R. F. & Atkins, J. F. P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA* **10**, 221–230 (2004).
- Horsfield, J. A., Wilson, D. N., Mannering, S. A., Adamski, F. M. & Tate, W. P. Prokaryotic ribosomes recode the HIV-1 *gag-pol*-1 frameshift sequence by an E/P site post-translocation simultaneous slippage mechanism. *Nucleic Acids Res.* **23**, 1487–1494 (1995).
- Lopinski, J. D., Dinman, J. D. & Bruenn, J. A. Kinetics of ribosomal pausing during programmed –1 translational frameshifting. *Mol. Cell. Biol.* **20**, 1095–1103 (2000).
- Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. *Cell* **55**, 447–458 (1988).
- Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).
- Weiss, R. B., Dunn, D. M., Shuh, M., Atkins, J. F. & Gesteland, R. F. *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.* **1**, 159–169 (1989).
- Léger, M., Dulude, D., Steinberg, S. V. & Brakier-Gingras, L. The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed –1 ribosomal frameshift. *Nucleic Acids Res.* **35**, 5581–5592 (2007).
- Larsen, B., Gesteland, R. F. & Atkins, J. F. Structural probing and mutagenic analysis of the stem-loop required for *Escherichia coli* *dnaX* ribosomal frameshifting: programmed efficiency of 50%. *J. Mol. Biol.* **271**, 47–60 (1997).
- Larsen, B., Wills, N. M., Gesteland, R. F. & Atkins, J. F. rRNA–mRNA base pairing stimulates a programmed –1 ribosomal frameshift. *J. Bacteriol.* **176**, 6842–6851 (1994).
- Chen, J. *et al.* High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proc. Natl Acad. Sci. USA* **111**, 664–669 (2014).
- Kim, H. K. *et al.* A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc. Natl Acad. Sci. USA* **111**, 5538–5543 (2014).
- Chen, J., Petrov, A., Tsai, A., O'Leary, S. E. & Puglisi, J. D. Coordinated conformational and compositional dynamics drive ribosome translocation. *Nature Struct. Mol. Biol.* **20**, 718–727 (2013).
- Chen, J., Tsai, A., Petrov, A. & Puglisi, J. D. Nonfluorescent quenchers to correlate single-molecule conformational and compositional dynamics. *J. Am. Chem. Soc.* **134**, 5734–5737 (2012).
- Aitken, C. E. & Puglisi, J. D. Following the intersubunit conformation of the ribosome during translation in real time. *Nature Struct. Mol. Biol.* **17**, 793–800 (2010).
- Tsuchihashi, Z. & Brown, P. O. Sequence requirements for efficient translational frameshifting in the *Escherichia coli* *dnaX* gene and the role of an unstable interaction between tRNA<sup>lys</sup> and an AAG lysine codon. *Genes Dev.* **6**, 511–519 (1992).
- Bertrand, C., Prere, M. F., Gesteland, R. F., Atkins, J. F. & Fayet, O. Influence of the stacking potential of the base 3' of tandem shift codons on –1 ribosomal frameshifting used for gene expression. *RNA* **8**, 16–28 (2002).
- Johansson, M., Zhang, J. & Ehrenberg, M. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc. Natl Acad. Sci. USA* **109**, 131–136 (2012).
- Tourigny, D. S., Fernandez, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, 1235490 (2013).
- Hughes, D., Atkins, J. F. & Thompson, S. Mutants of elongation factor Tu promote ribosomal frameshifting and nonsense readthrough. *EMBO J.* **6**, 4235–4239 (1987).
- Qu, X. *et al.* The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475**, 118–121 (2011).
- Liu, C. Y., Qureshi, M. T. & Lee, T. H. Interaction strengths between the ribosome and tRNA at various steps of translocation. *Biophys. J.* **100**, 2201–2208 (2011).
- Valle, M. *et al.* Locking and unlocking of ribosomal motions. *Cell* **114**, 123–134 (2003).
- Qin, P., Yu, D., Zuo, X. & Cornish, P. V. Structured mRNA induces the ribosome into a hyper-rotated state. *EMBO Rep.* **15**, 185–190 (2014).
- Jacks, T. *et al.* Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* **331**, 280–283 (1988).
- Liao, P. Y., Choi, Y. S., Dinman, J. D. & Lee, K. H. The many paths to frameshifting: kinetic modelling and analysis of the effects of different elongation steps on programmed –1 ribosomal frameshifting. *Nucleic Acids Res.* **39**, 300–312 (2011).
- Uemura, S. *et al.* Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012–1017 (2010).

**Acknowledgements** This work was supported by US National Institutes of Health (NIH) grant GM51266 to J.C., A.T. and J.D.P.; by NIH grant GM099687 to A.P., S.E.O'L. and J.D.P.; Wenner-Gren Foundations (Stockholm) to M.J.; and by a Stanford Interdisciplinary Graduate Fellowship to J.C. The authors thank D. Hsu and R. Dalal (Pacific Biosciences Inc.) for their assistance on the ZMW instrumentation; and S. Yan, H. K. Kim and I. Tinoco Jr (University of California at Berkeley) for helpful discussions. J.C. would like to thank I. Lin for support.

**Author Contributions** J.C. performed all the experiments and the data analysis. J.C., A.P., A.T. and J.D.P. designed the project and wrote the manuscript. M.J. and S.E.O'L. assisted with reagent preparation. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.P. ([puglisi@stanford.edu](mailto:puglisi@stanford.edu)).

# X-ray structures of GluCl in *apo* states reveal a gating mechanism of Cys-loop receptors

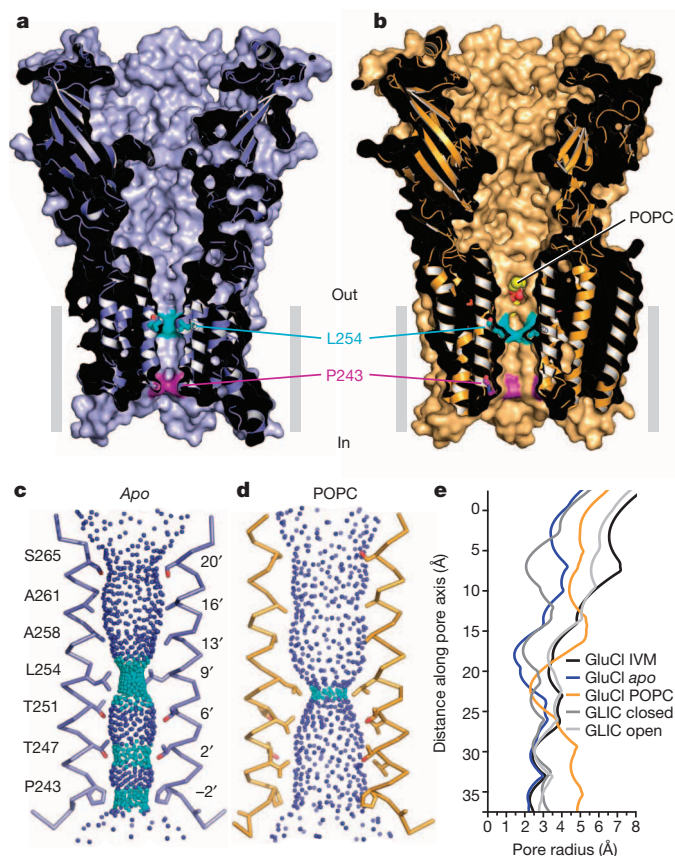
Thorsten Althoff<sup>1†\*</sup>, Ryan E. Hibbs<sup>1†\*</sup>, Surajit Banerjee<sup>2</sup> & Eric Gouaux<sup>1,3</sup>

Cys-loop receptors are neurotransmitter-gated ion channels that are essential mediators of fast chemical neurotransmission and are associated with a large number of neurological diseases and disorders, as well as parasitic infections<sup>1–4</sup>. Members of this ion channel superfamily mediate excitatory or inhibitory neurotransmission depending on their ligand and ion selectivity. Structural information for Cys-loop receptors comes from several sources including electron microscopic studies of the nicotinic acetylcholine receptor<sup>5</sup>, high-resolution X-ray structures of extracellular domains<sup>6</sup> and X-ray structures of bacterial orthologues<sup>7–10</sup>. In 2011 our group published structures of the *Caenorhabditis elegans* glutamate-gated chloride channel (GluCl) in complex with the allosteric partial agonist ivermectin, which provided insights into the structure of a possibly open state of a eukaryotic Cys-loop receptor, the basis for anion selectivity and channel block, and the mechanism by which ivermectin and related molecules stabilize the open state and potentiate neurotransmitter binding<sup>11</sup>. However, there remain unanswered questions about the mechanism of channel opening and closing, the location and nature of the shut ion channel gate, the transitions between the closed/resting, open/activated and closed/desensitized states, and the mechanism by which conformational changes are coupled between the extracellular, orthosteric agonist binding domain and the transmembrane, ion channel domain. Here we present two conformationally distinct structures of *C. elegans* GluCl in the absence of ivermectin. Structural comparisons reveal a quaternary activation mechanism arising from rigid-body movements between the extracellular and transmembrane domains and a mechanism for modulation of the receptor by phospholipids.

We obtained three-dimensional crystals of GluCl in the absence of ivermectin by supplementing the previously characterized receptor–Fab complex<sup>11</sup> with 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and either setting up crystallizations immediately or after a 4-week incubation, which yielded an *apo* state or a POPC-bound conformation, respectively. The structures of GluCl in an *apo* state or in complex with POPC show a solvent-accessible pathway from the outermost region of the extracellular domain, through the vase-shaped extracellular vestibule, to the transmembrane ion channel pore (Fig. 1a, b, Extended Data Fig. 1 and Extended Data Table 1). The ion channel pore is lined by the M2 transmembrane helices, with Pro 243 and Leu 254 occupying key sites at the cytoplasmic and middle portion of the ion channel. In the POPC complex, we visualized lipid molecules bound between subunits, near the extracellular side of the transmembrane domain, with their head groups wedged between the M1 and M3 helices of adjacent subunits (Fig. 1b and Extended Data Fig. 1c, d).

In the *apo* state the M2 helices are nearly parallel to the pore axis with three narrow regions at Pro 243, Thr 247 and Leu 254 (Fig. 1c). The pore is most constricted at Leu 254 (Leu 9' on the M2, pore-lining helix), with the hydrophobic side chain of the leucine residue restricting the pore radius to ~1.4 Å, too small for the conduction of chloride ions,

suggesting that Leu 254 forms the shut gate of the ion channel pore (Extended Data Fig. 2)<sup>12,13</sup>. Previous observations that mutation of Leu 9' perturbs ion channel gating are in harmony with the hypothesis that Leu 9' has an important role in channel function<sup>14–16</sup>. We suggest that



**Figure 1 | Apo and POPC-bound GluCl.** **a**, Sagittal slice along the pore axis of the *apo* GluCl structure showing the solvent-accessible surface and underlying secondary structure. **b**, Sagittal slice through the pore of GluCl in complex with POPC, similar to panel **a**. Atoms of the POPC head group are visible through a fenestration between adjacent subunits. **c**, Solvent contours of the transmembrane pore of the *apo* GluCl pore showing the M2 helices of subunits P and R and the side chains of pore-lining residues, numbered according to protein sequence and position in the M2 helix. Small blue spheres define a radius >2.8 Å and cyan spheres represent a radius of 1.4–2.8 Å. **d**, Contours of the POPC-bound pore, similar to panel **c**. **e**, Illustration of the pore radii as a function of distance along the pore axis for *apo*, POPC- and ivermectin-bound GluCl, along with the open and closed states of GLIC (Protein Data Bank accessions 3EAM and 4NPQ). Pore radii in panels **c–e** were calculated using the computer program Hole. IVM, ivermectin.

<sup>1</sup>Vollum Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA. <sup>2</sup>NE-CAT/Cornell University, 9700 South Cass Avenue, Building 436 E001, Argonne, Illinois 60439, USA. <sup>3</sup>Howard Hughes Medical Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA. <sup>†</sup>Present addresses: Department of Physiology, David Geffen School of Medicine, University of California Los Angeles, 10833 Le Conte Avenue, Los Angeles, California 90095-1751, USA (T.A.); Department of Neuroscience, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390-9111, USA (R.E.H.).

\*These authors contributed equally to this work.

this *apo* structure of GluCl defines the closed/resting state of this eukaryotic Cys-loop receptor.

In the POPC-bound structure, the ion channel pore is also straight, yet wider than in the *apo* state, with a constriction at Leu 254 yielding a pore radius of  $\sim 2.4$  Å (Fig. 1d), similar in size to the narrow region of the pore in the ivermectin-bound state. Comparisons of the pore radii of the *apo* and ivermectin-bound states of GluCl with ELIC (*Erwinia chrysanthemi* ligand-gated ion channel) and GLIC (*Gloeobacter violaceus* ligand-gated ion channel)<sup>7–10</sup> show how the dimensions are remarkably similar near the cytoplasmic region of the pore, yet diverge substantially at the extracellular entrance (Fig. 1e and Extended Data Fig. 3). While the position of the shut gate in the ‘closed’ GLIC structure is similar to that of *apo* GluCl, the pore of GLIC is narrower in comparison to the pore of GluCl in the area extracellular to the gate. Additional studies are required to define the ion-conducting properties of both the ivermectin- and POPC-bound states of GluCl.

Inspection of electron density maps derived from GluCl crystals grown in the presence of POPC after an  $\sim 4$ -week incubation with lipid revealed prominent densities located between transmembrane segments M1 and M3 of adjacent subunits in 8 of the 10 subunit interfaces in the asymmetric unit (Extended Data Fig. 4a, b). We modelled these densities as POPC molecules with the phosphocholine head group pointing towards the centre of the pore and the two alkyl tails located on the periphery of the transmembrane domain (Fig. 2a, b and Extended Data Fig. 4a, b). The region occupied by POPC molecules overlaps with the ivermectin site derived from the GluCl–ivermectin complex (Fig. 2b)<sup>11</sup>, a binding pocket that recent molecular dynamics simulations also identified as an intersubunit crevice transiently occupied by, on average, four lipid molecules per pentamer<sup>17</sup>.

We screened several lipids in binding assays, finding that 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoserine (POPS) competes for ivermectin binding with an inhibition constant ( $K_i$ ) of  $\sim 167$   $\mu$ M (Fig. 2c and Extended Data Fig. 4c, e). Unfortunately POPS, while binding strongly, does not

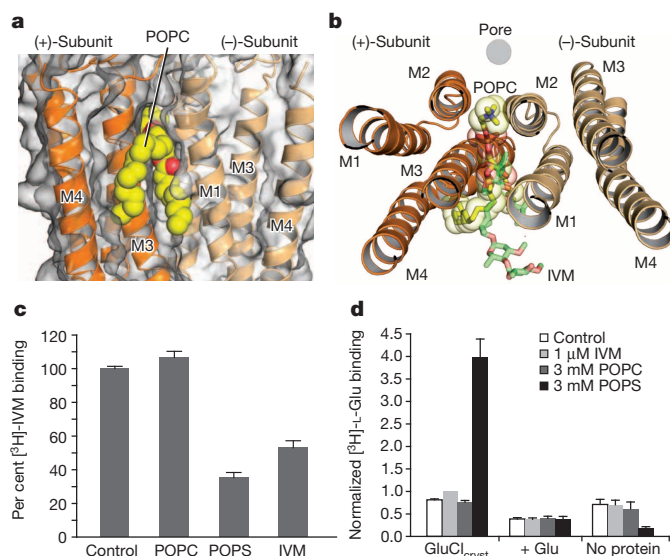
yield well-diffracting crystals whereas POPC binds weakly and does not measurably compete for ivermectin binding. We therefore used POPS in binding experiments and POPC in the structural studies. POPS potentiates glutamate binding (Fig. 2d), like ivermectin (dissociation constant ( $K_d$ ) for L-glutamate in presence of ivermectin is  $\sim 0.66$   $\mu$ M)<sup>11</sup>, yielding a  $K_d$  for glutamate binding of  $\sim 1.1$   $\mu$ M (Extended Data Fig. 4d). Thus, ivermectin or lipids can occupy the intersubunit crevice within the membrane-spanning region of the receptor and potentiate neurotransmitter binding, providing insight into the long-standing observation of small molecule and lipid modulation of agonist binding in Cys-loop receptors<sup>18,19</sup>.

Superpositions of individual subunits from the *apo*, POPC- and ivermectin-bound states demonstrate that the extracellular and transmembrane domains move largely as rigid bodies, undergoing movements relative to one another (Extended Data Fig. 5). Thus, superposition of the extracellular domain from a single subunit of the *apo* and ivermectin-bound structures shows that, during activation by ivermectin, the transmembrane domain undergoes a screw-axis-like movement, rotating around an axis tipped about  $40^\circ$  off the pore axis and shifting towards the extracellular side of the membrane by  $\sim 4.5$  Å. Transition from the *apo* to the ivermectin-bound state thereby involves tilting of the pore-lining M2 helix by  $\sim 8^\circ$  ‘away’ from the ion channel, which relieves the occlusion of the pore by Leu 254 (Fig. 3a and Extended Data Fig. 6).

The conformations stabilized by ivermectin and by POPC are markedly different. In comparing the *apo* and POPC-bound states, the transmembrane domain undergoes a rotation about an axis approximately parallel to the pore, which in turn gives rise to a displacement by  $\sim 3$  Å ‘away’ from the pore axis in the plane of the membrane. Together these movements lead to an expansion of the ion channel pore while the M2 helices remain oriented parallel to the pore axis (Fig. 3b). A comparison of the ivermectin- and POPC-bound states shows a large relative movement of the extracellular and transmembrane domains, with the M2 helix undergoing a tilt by  $\sim 8.7^\circ$  and a translational movement of  $\sim 6$  Å (Fig. 3c). Participating in these relative movements of the extracellular and transmembrane domains is the namesake Cys-loop, which is cradled in a concave depression on ‘top’ of the transmembrane helices, stabilized by interactions that include the  $\beta 10$ –M1 covalent connection and the interface between the  $\beta 1$ – $\beta 2$  and M2–M3 loops (Fig. 3a–c)<sup>5,20</sup>.

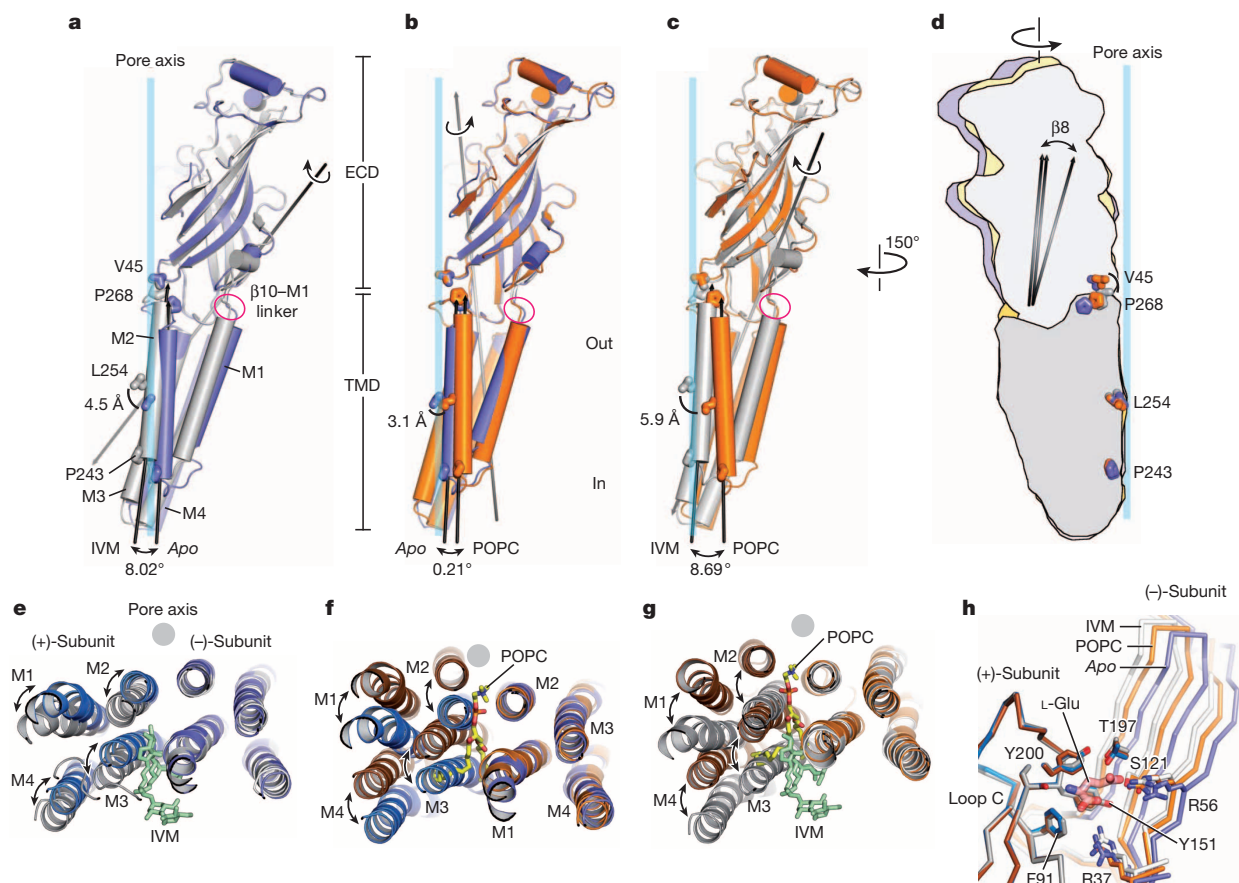
Superpositions of the transmembrane domains of the *apo*, POPC- and ivermectin-bound structures further illustrate the relative conformational changes between the extracellular and transmembrane domains (Fig. 3d and Extended Data Fig. 6). Here we observe mainly two movements: the upper part of the extracellular domain, as marked by the  $\alpha 1$  helix, twists around the pore axis, and the lower part, as exemplified by the  $\beta 8$  strand, tilts towards the centre of the pore. For both movements, displacements are largest for the transition to the ivermectin-bound state. Loop C does not close the neurotransmitter binding site by an independent motion but closure is rather a consequence of the rigid-body twist of the extracellular domain of each subunit. While the observed closure of loop C is mechanistically distinct from that observed in AChBP (refs 21, 22), it is consistent with biochemical studies of nicotinic receptors wherein agonist binding results in protection of the tip of loop C from reducing reagents<sup>23</sup>.

To visualize the conformational changes associated with ivermectin binding, we superimposed the transmembrane domains of the (–)-subunits for the *apo* and ivermectin states (Extended Data Fig. 7). Inspection of the transmembrane helices in the (+)-subunit shows that in the *apo* state the space between the M3 and M1 helices of the (+)- and the (–)-subunits is ‘collapsed’ (Fig. 3e). When ivermectin inserts into this site, helices M1–M4 undergo an anticlockwise rotation of  $\sim 10^\circ$  relative to the pore axis, ‘splaying open’ the intersubunit interface. This movement increases the  $\alpha$ -carbon distance between Leu 218 (M1) and Gly 281 (M3), the latter of which is crucial for ivermectin sensitivity<sup>24</sup>, from  $\sim 6.9$  Å in the *apo* state to  $\sim 9.3$  Å in the ivermectin complex. The M2–M3 loop not only participates in direct contacts with ivermectin via Ile 273 (ref. 11), but it also connects helix M3 of the ivermectin site



**Figure 2 | Phospholipids occupy intersubunit site, compete with ivermectin and potentiate glutamate binding.** **a**, **b**, POPC binding site between M1 and M3 helices of adjacent subunits viewed parallel (**a**) and perpendicular (**b**) to the membrane. In **b** we show the location of the pore by a grey circle and the overlap between POPC and ivermectin (shown in ‘sticks’ representation). **c**, Radioligand competition experiment using GluCl and [<sup>3</sup>H]-ivermectin (control) and cold POPC, POPS or ivermectin. **d**, POPS potentiates glutamate binding as demonstrated by a [<sup>3</sup>H]-L-glutamate binding experiment in the presence of POPC, POPS or ivermectin. Bars are normalized to the extent of binding in the presence of ivermectin. For panels **c**, **d**, experiments were carried out three separate times, with experiments done in triplicate. Points are mean values and error bars represent s.e.m.





**Figure 3 | Conformational changes of extracellular and transmembrane domains.** **a–c**, Superpositions of residues 1–211 of the extracellular domains illustrate a screw-axis-like conformational change within GluCl subunits. **a–c**, Superposition of *apo* (blue) and iверmectin (grey) subunits (**a**), *apo* and POPC (orange) subunits (**b**) and POPC and iверmectin subunits (**c**). Panel **d** illustrates relative conformational changes within subunits when the transmembrane domains are superposed, using residues 212–342, and the same colour coding as in panels **a–c**. **e**, Superposition of transmembrane domains of the (-)-subunits for the *apo* and iверmectin structures, illustrating relative movements of (+) transmembrane domain. View is from the

extracellular side. **f, g**, Similar superpositions as in panel **e** for the *apo* and POPC structures (**f**) and for the POPC and iверmectin structures (**g**). **h**, Neurotransmitter binding site is more open in the *apo* state in comparison to the POPC and iверmectin states. Superposition of residues in the extracellular domain of the (+)-subunit illustrates the relative displacement of residues contributing to the neurotransmitter binding site on the (-)-subunit, including Arg 37 and Arg 56, in the *apo* (blue), POPC (orange) and iверmectin (grey) states. The POPC-bound state represents an intermediate position between the *apo* and iверmectin-bound states.

with helix M2 of the ion channel pore, thus providing a direct coupling for the binding of iверmectin with the tilting of the M2 helix  $\sim 4$  Å away from the five-fold axis and opening of the pore.

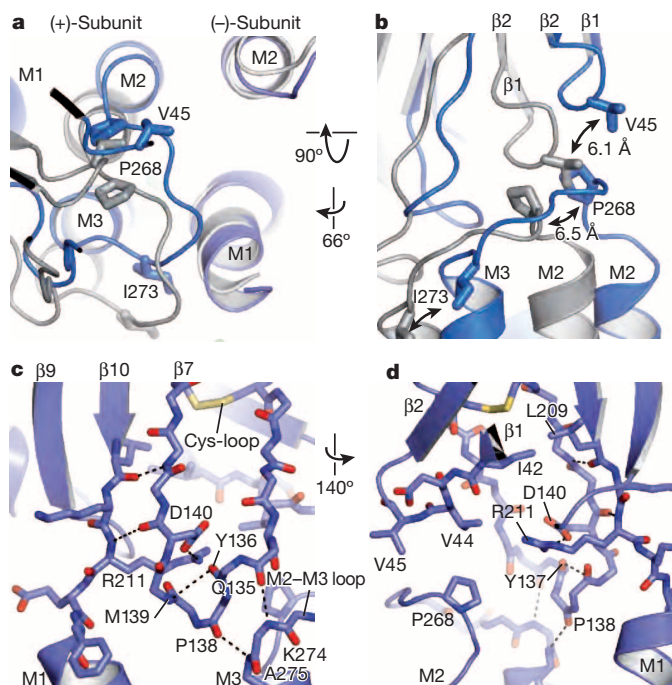
In comparison to iверmectin, the longer POPC head group inserts deeply into the intersubunit crevice. Thus, analysis of the POPC complex relative to the *apo* state shows that transmembrane segments of the (+)-subunit undergo a greater displacement in comparison to the iверmectin complex, with the (+) transmembrane bundle moving  $\sim 5.7$  Å towards the ion channel pore (Fig. 3f and Extended Data Fig. 7). Nevertheless, the separation between subunits is similar to the iверmectin complex, as measured by the 9.4 Å distance between the  $\alpha$ -carbons of residues Leu 218 (M1; (-)-subunit) and Gly 281 (M3; (+)-subunit).

A remarkable plasticity of the transmembrane domains is demonstrated by comparison of the iверmectin and POPC complexes (Extended Data Fig. 7). Here, superposition of the transmembrane regions of the (-)-subunit shows that the transmembrane bundle of the (+)-subunit from the POPC complex moves  $\sim 8.8$  Å 'away' from the ion channel pore. This shift of nearly the diameter of an  $\alpha$ -helix results in the (+) M3 helix occupying the position of the (+) M2 helix in the iверmectin complex, thus showing how, in these two complexes, M2 replaces M3 at the interface with M2 of the (-)-subunit (Fig. 3g).

To analyse the changes at the orthosteric glutamate binding site we superimposed the extracellular domains of the (+)-subunit, a facet of the pocket that harbours multiple elements of the agonist binding site,

including an aromatic box closed by loop C (refs 6, 11, 21). Following this superposition, we see that the  $\beta$ -strands on the (-)-subunit shift closer to the (+)-subunit in the POPC and iверmectin complexes, with the largest shift seen in the iверmectin-bound state (Fig. 3h). These shifts close the binding pocket, moving key residues towards the (+)-subunit and thereby, we speculate, strengthening neurotransmitter binding. Consistent with this notion is our earlier observation that, like iверmectin, POPC binding potentiates binding of glutamate. Indeed, in the iверmectin-bound conformation the  $\alpha$ -carbon atoms of Ser 121 and Arg 56 move by 2.2 Å and 2.5 Å, respectively, from their positions in the *apo* state (Fig. 3h). The positions of neurotransmitter-binding residues in the POPC complex are intermediate between their respective positions in the *apo* and iверmectin complex, thus providing an explanation of how lipids, such as POPs, might potentiate neurotransmitter binding.

The extracellular and transmembrane domains are covalently connected by the  $\beta$ 10–M1 linker yet they also interact via contacts between the  $\beta$ 1– $\beta$ 2 loop and the M2–M3 loop as well as contacts between the Cys-loop and the extracellular ends of the transmembrane helices. Indeed, the M2–M3 linker, the  $\beta$ 1– $\beta$ 2 loop and the Cys-loop are 'hotspots' of non-covalent bonds and steric interactions that have been extensively studied<sup>20,25,26</sup>. During channel opening, as defined by the transition from the *apo* to the iверmectin-bound state, the M2–M3 loop shifts by more than 5 Å away from the ion channel pore, as visualized by the movement of Pro 268 of the M2–M3 loop passing beneath Val 45 on the  $\beta$ 1– $\beta$ 2

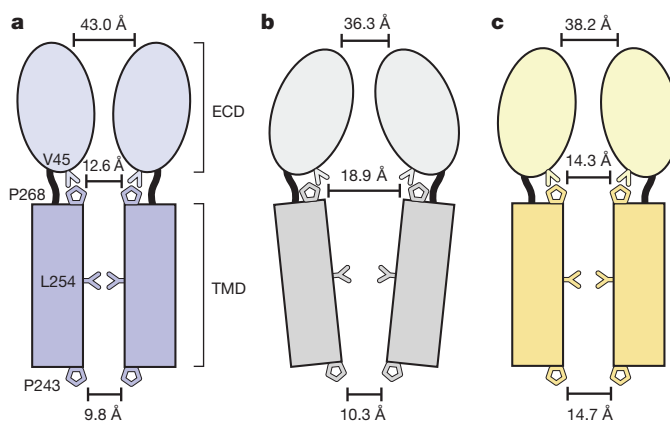


**Figure 4 | The M2–M3 loop couples conformational changes between the transmembrane and extracellular domains.** **a**, Superposition of the transmembrane region of the (–)-subunits for the *apo* (blue) and ivermectin (grey) states illustrates the relative movement of the transmembrane domains and the M2–M3 loop of the neighbouring (+)-subunit. **b**, Same superposition as in panel **a**, viewed approximately parallel to the membrane, showing the coupled movement of the  $\beta$ 1– $\beta$ 2 loop and the M2–M3 loop, emphasizing key residues Val 45, Pro 268 and Ile 273. **c**, **d**, Illustration of key residues forming hydrogen bonds and a salt bridge that connect the transmembrane and extracellular domains in the *apo* structure seen from two directions approximately parallel to the membrane. No salt bridge is formed between the pre-M1 linker and the  $\beta$ 1– $\beta$ 2 loop (**d**).

loop (Fig. 4a, b). In the ivermectin-bound conformation, Val 45 is lodged against Pro 268, thus providing a steric block on the M2–M3 loop and, in turn, stabilizing the M2 and M3 helices and the entire transmembrane domain in an open pore conformation<sup>27</sup>. Furthermore, Pro 268 is strictly conserved throughout the family of Cys-loop receptors, and mutations of this residue, as well as others nearby, have profound effects on the channel gating and desensitization behaviour<sup>1,28</sup>.

The extracellular end of the M3 helix interacts with the Cys-loop via a hydrogen bond between the carbonyl oxygens of Gln 135 and Pro 138 and the amide nitrogens of Lys 274 and Ala 275, respectively (Fig. 4c). The Cys-loop itself is also stabilized by a backbone hydrogen bond between Tyr 136 and Met 139 and is coupled to the extracellular end of the M1 helix by a hydrogen bond between backbone atoms of Asp 140 and Arg 211, a salt bridge between the side chains of Arg 211 and Asp 140 and the carbonyl oxygen of Tyr 137, as well as interactions between strands  $\beta$ 7 and  $\beta$ 10 (Fig. 4c, d)<sup>29</sup>. All these interactions remain intact in the three GluCl conformations (Extended Data Fig. 8). Through this route movements of the M1 helix can be directly transmitted to the M3 helix in the same subunit. Additionally, changes in the transmembrane domain can be transmitted to the (+) side of the ligand binding site via the  $\beta$ 7 and  $\beta$ 10 strands. Nevertheless, we do not find evidence for a direct coupling between the  $\beta$ 10–M1 linker and  $\beta$ 1– $\beta$ 2 loop as in the bacterial orthologues<sup>7</sup> and in the acetylcholine receptor<sup>30</sup>. Because the charged amino acids in the  $\beta$ 1– $\beta$ 2 loop are replaced by Val 44 in GluCl, no salt bridge with Arg 211 can be formed (Fig. 4d).

Global superpositions of pentamers from the *apo*, ivermectin and POPC-bound states illustrate quaternary conformational differences (Fig. 5, Extended Data Fig. 9 and Supplementary Movies 1 and 2). In the *apo* state the extracellular domains are separated by ~43 Å as measured



**Figure 5 | Conformational changes in the pentamer.** **a–c**, Schematic illustration of the conformations of the *apo*-closed (**a**), ivermectin (**b**) and POPC-bound (**c**) states as seen in the plane of the membrane.

at Thr 11 in helix 1 in two opposing subunits. The M2 helices are straight and oriented perpendicular to the plane of the membrane with distances of 12.6 Å at Ser 265 (pore apex) and 9.8 Å at Pro 243 (pore base) and the pore is occluded by the side chains of Leu 254. In the ivermectin- and POPC-bound states the upper parts of the extracellular domains tilt towards the pore in a motion resembling the closure of a blossom. The distances at helix 1 shrink to 36.3 Å in the ivermectin-bound state and to 38.2 Å in the POPC-bound conformation (Extended Data Fig. 9a–c). The changes in the transmembrane domain are markedly different between the ivermectin- and the lipid-bound conformations. In the former case the M2 helices tilt away from the five-fold axis and the distance at the pore apex increases to 18.9 Å while it remains nearly constant at the intracellular side. Helices M3 and M4 rotate clockwise around the centre of the helix bundle resulting in an apparent overall clockwise rotation of the whole transmembrane domain of the receptor. In contrast, during the transition from the *apo* to the POPC-bound state the M2 helices undergo a clockwise twist around the pore axis increasing the distance to 14.3 Å and 14.7 Å (pore apex and base), thus remaining straight. This quaternary structural change leads to an iris-like opening of the pore and causes M1 and M3 to be displaced in an anticlockwise rotation with little displacement of M4 (Extended Data Fig. 9d–f).

We hypothesize that the structures of GluCl in the *apo* and ivermectin-bound forms represent the closed/resting state and a potentially open/activated state of a eukaryotic Cys-loop receptor, respectively, showing how the shut ion channel gate is defined by a hydrophobic belt of five leucine residues and how the possible opening of the ion channel pore involves the outward tilting of approximately straight M2 transmembrane helices and the inward contraction of the extracellular domain (Fig. 5a–c). Accompanying these quaternary changes is a ‘bend/twist’ of each subunit at the junction of the extracellular–transmembrane domain boundary, where the namesake Cys-loop acts like a ball in the socket of the extracellular end of the transmembrane domain, interactions further stabilized by the  $\beta$ 10–M1 connection and by interactions between the M2–M3 and  $\beta$ 1– $\beta$ 2 loops. The GluCl–POPC complex demonstrates how lipids can allosterically modulate Cys-loop receptor function, inducing an expanded, open-like conformation of the transmembrane domain and potentiating agonist binding. Taken together, these studies provide motivation for future experiments and insight into the gating, modulation and structural plasticity of eukaryotic Cys-loop receptors.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 March; accepted 10 July 2014.

1. Thompson, A. J., Lester, H. A. & Lummis, S. C. The structural basis of function in Cys-loop receptors. *Q. Rev. Biophys.* **43**, 449–499 (2010).

2. Sine, S. M. End-plate acetylcholine receptor: structure, mechanism, pharmacology, and disease. *Physiol. Rev.* **92**, 1189–1234 (2012).
3. Corringer, P. J. *et al.* Structure and pharmacology of pentameric receptor channels: from bacteria to brain. *Structure* **20**, 941–956 (2012).
4. Boatin, B. A. & Richards, F. O. J. Control of onchocerciasis. *Adv. Parasitol.* **61**, 349–394 (2006).
5. Unwin, N. Refined structure of the nicotinic acetylcholine receptor. *J. Mol. Biol.* **346**, 967–989 (2005).
6. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
7. Hilf, R. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457**, 115–118 (2009).
8. Hilf, R. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
9. Bocquet, N. *et al.* X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* **457**, 111–114 (2009).
10. Sauguet, L. *et al.* Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation. *Proc. Natl Acad. Sci. USA* **111**, 966–971 (2014).
11. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* **474**, 54–60 (2011).
12. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
13. Bali, M. & Akabas, M. H. The location of a closed channel gate in the GABBA receptor channel. *J. Gen. Physiol.* **129**, 145–159 (2007).
14. Revah, F. *et al.* Mutations in the channel domain alter desensitization of a neuronal nicotinic receptor. *Nature* **353**, 846–849 (1991).
15. Labarca, C. *et al.* Channel gating governed symmetrically by conserved leucine residues in the M2 domain of nicotinic receptors. *Nature* **376**, 514–516 (1995).
16. Filatov, G. N. & White, M. M. The role of conserved leucines in the M2 domain of the acetylcholine receptor in channel gating. *Mol. Pharmacol.* **48**, 379–384 (1995).
17. Yolk, O., Brömstrup, T., Bertaccini, E. J., Trudell, J. R. & Lindahl, E. Stabilization of the GluCl ligand-gated ion channel in the presence and absence of ivermectin. *Biophys. J.* **105**, 640–647 (2013).
18. Barrantes, F. J. Structural basis for lipid modulation of nicotinic acetylcholine receptor function. *Brain Res. Brain Res. Rev.* **47**, 71–95 (2004).
19. daCosta, C. J., Dey, L., Therien, J. P. & Baenziger, J. E. A distinct mechanism for activating uncoupled nicotinic acetylcholine receptors. *Nature Chem. Biol.* **9**, 701–707 (2013).
20. Lee, W. Y., Free, C. R. & Sine, S. M. Nicotinic receptor interloop proline anchors beta1-beta2 and Cys loops in coupling agonist binding to channel gating. *J. Gen. Physiol.* **132**, 265–278 (2008).
21. Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
22. Hansen, S. B. *et al.* Structures of Aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24**, 3635–3646 (2005).
23. Darnley, V. N. & Karlin, A. Effects of agonists and antagonists on the reactivity of the binding site disulfide in acetylcholine receptor from *Torpedo californica*. *Biochemistry* **19**, 3924–3932 (1980).
24. Lynagh, T. & Lynch, J. W. A glycine residue essential for high ivermectin sensitivity in Cys-loop ion channel receptors. *Int. J. Parasitol.* **40**, 1477–1481 (2010).
25. Velisetty, P., Chalamalasetti, S. V. & Chakrapani, S. Structural basis for allosteric coupling at the membrane-protein interface in *Gloeobacter violaceus* ligand-gated ion channel (GLIC). *J. Biol. Chem.* **289**, 3013–3025 (2014).
26. Reeves, D. C., Jansen, M., Bali, M., Lemster, T. & Akabas, M. H. A role for the  $\beta$ 1- $\beta$ 2 loop in the gating of 5HT3 receptors. *J. Neurosci.* **25**, 9358–9366 (2005).
27. Calimet, N. *et al.* A gating mechanism of pentameric ligand-gated ion channels. *Proc. Natl Acad. Sci. USA* **110**, E3987–E3996 (2013).
28. Purohit, P., Gupta, S., Jadey, S. & Auerbach, A. Functional anatomy of an allosteric protein. *Nature Commun.* **4**, 2984 (2013).
29. Gleitsman, K. R., Lester, H. A. & Dougherty, D. A. Probing the role of backbone hydrogen bonding in a critical beta sheet of the extracellular domain of a cys-loop receptor. *ChemBioChem* **10**, 1385–1391 (2009).
30. Lee, W. Y. & Sine, S. M. Principle pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank all staff of beamline 24-ID-C at the Advanced Photon Source. We thank L. Vaskalis and H. Owen for help in figure and manuscript preparation, respectively. D. Cawley at the Vaccine and Gene Therapy Institute, OHSU, provided the monoclonal antibody. We appreciate discussions with Gouaux laboratory members. This work was supported by a postdoctoral fellowship (Forschungstipendium AL 1725-1/1) from the Deutsche Forschungsgemeinschaft to T.A. and an individual NIH National Research Service Award (F32NS061404) to R.E.H.; E.G. is supported by the NIH and is an investigator with the Howard Hughes Medical Institute.

**Author Contributions** T.A., R.E.H. and S.B. performed the experiments and T.A., R.E.H. and E.G. wrote the manuscript.

**Author Information** The coordinates and structure factors for the GluCl *apo* and POPC-bound structures have been deposited in the Protein Data Bank under accession codes 4TNV and 4TNW, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. ([gouauxe@ohsu.edu](mailto:gouauxe@ohsu.edu)).



## CORRIGENDUM

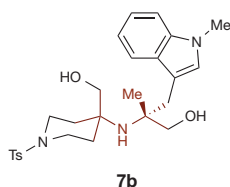
doi:10.1038/nature13635

### Corrigendum: Palladium-catalysed C–H activation of aliphatic amines to give strained nitrogen heterocycles

Andrew McNally, Benjamin Haffemayer, Beatrice S. L. Collins & Matthew J. Gaunt

*Nature* **510**, 129–133 (2014); doi:10.1038/nature13389

The structure of compound **7b** in Fig. 3e of this Letter has been corrected in Fig. 1 of this Corrigendum.



**Figure 1** | This figure shows the corrected structure of compound **7b** in Fig. 3e of the original Letter.

## RETRACTION

doi:10.1038/nature13661

### Retraction: Generation of pluripotent stem cells from adult human testis

Sabine Conrad, Markus Renninger, Jörg Hennenlotter, Tina Wiesner, Lothar Just, Michael Bonin, Wilhelm Aicher, Hans-Jörg Bühring, Ulrich Mattheus, Andreas Mack, Hans-Joachim Wagner, Stephen Minger, Matthias Matzkies, Michael Reppel, Jürgen Hescheler, Karl-Dietrich Sievert, Arnulf Stenzl & Thomas Skutella

*Nature* **456**, 344–349 (2008); doi:10.1038/nature07404

corrigendum *Nature* **460**, 1044 (2009); doi:10.1038/nature08353

The authors have provided new data to correct errors presented in this Article. *Nature* has peer-reviewed all evidence provided by the authors to the editors. The images presented in the original version of the Article made the data appear more robust than newly conducted experiments show. The new data have brought to light that the original conclusions are not as robust as presented in the original paper. *Nature* does not dispute the main claim that the cells are pluripotent to some level, but the level of proof of pluripotency shown is not in line with regular criteria for such papers in *Nature*. Consequently, the authors have agreed to retract their manuscript.

**Correspondence** should be addressed to T.S. (skutella@ana.uni-heidelberg.de).

# CAREERS

@NATUREJOBS Follow us on Twitter for the latest on science jobs [twitter.com/naturejobs](https://twitter.com/naturejobs)

NATUREJOBS BLOG Discussions and tips on science careers [blogs.nature.com/naturejobs](https://blogs.nature.com/naturejobs)

NATUREJOBS For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

JOSE FLUSTE RAGA/CORBIS



Houston is hiring: Rice University and the University of Houston are among the institutions looking to expand their staff thanks to generous funding.

## BIOMEDICAL SCIENCE

# Houston has lift-off

*Buoyed by state funding, biomedical sciences are booming in the Texan city.*

BY PAUL SMAGLIK

**T**exans pride themselves on their state's massive scale. So it is no surprise that Houston, the Lone Star State's most populous city and the fourth largest in the United States, boasts a biomedical infrastructure of impressive proportions.

What is surprising, however, is the sheer density of that infrastructure. The Texas Medical Center (TMC) is essentially a city of hospitals, clinics and medical-research labs — 59 institutions and organizations in all — jammed into 544 hectares in southern Houston. Walking through what staff jokingly call “the world's largest parking garage” (parking fees help to pay for the operating costs) feels like strolling

through any dense urban centre — except that every building is a hospital, clinic, university or research institution.

That size, plus a concentrated infrastructure, serves as a magnet for patients, funding and, increasingly, scientific recruitment. Houston's hospitals serve more than 6 million patients a year, and the University of Texas MD Anderson Cancer Center leads the nation in grant dollars won from the US National Cancer Institute in Bethesda, Maryland. The city claims US\$420 million from the \$3-billion Cancer Prevention Research Institute of Texas (CPRIT), which since its inception has awarded 36 grants worth \$124 million for recruiting scientists, labs and companies to the city.

“What's unique is we have such a high density

of top-tier medical-research institutions in such a compact space,” says Ronald DePinho, president of MD Anderson. The mix of hospitals and academic institutions at the TMC provides a platform for growth in clinical research — such studies require a patient base and interdisciplinary research, which flourishes when neighbouring institutions collaborate. Nearby Rice University and the University of Texas provide a basic-research yin to the TMC's clinical yang, and bolster the city's interdisciplinary bona fides.

DePinho, civic leaders and academics are proud of the city's vast educational and clinical assets — the life and health sciences are the city's second biggest industry after energy and oil — but they feel the lack of a major biotechnology and pharmaceutical presence in ►



► **Houston.** The region has had a robust history of technology transfer in the life sciences, but has generated mostly licences, which produce revenue, rather than biotechnology companies, which create jobs. Houston also needs to train young scientists to fill positions in biotech subfields such as technology transfer and intellectual property, or else recruit talent from outside the area.

The city is beginning to address those needs. First, MD Anderson created the Center for Professional Development and Entrepreneurship in 2012, giving young scientists training options beyond tenure-track careers. Then, last year, the University of Texas Health Science Center began changing the way it prepares graduate students for a changing work environment (see 'Avenues beyond academia').

In addition, Robert Robbins, chief executive of the TMC, is leading a strategic plan to identify the region's 'connective tissue' in terms of cooperative interdisciplinary research. The plan currently focuses on five research themes: clinical research, genomics, health policy, innovation and regenerative medicine.

There are also tentative plans for a research facility that all TMC members could share, in the same way that the Broad Institute in Cambridge, Massachusetts, draws scientists from Harvard University and the Massachusetts Institute of Technology. "We need to pool our resources," Robbins says. If such a facility is built, it could mean hundreds of new positions, especially in the five key research themes.

### CULTURE OF SHARING

TMC members have a history of sharing academic resources, says George Stancel, an executive vice-president at the University of Texas Houston. The 'Green Sheet' — a list of seminars and talks on the TMC campus — dates back to the 1970s, when the calendar was printed on green paper. "Any day of the week I can walk into a seminar or meeting at another medical school, cancer centre, hospital or academic centre and learn from those talks and presentations," Stancel says. This year the TMC added 'The Lead' (short for Leadership, Education and Career Development), which is a similar service for professional development that includes workshops and presentations about non-academic careers.

Houston's biomedical institutions also tend to share when recruiting talent. The current chair of neurosurgery at the Baylor College of Medicine is also chair of neurosurgery at MD Anderson and director of MD Anderson's brain-tumour centre. The two institutions recruit jointly for this and other roles, which makes recruitment easier, especially when it involves bringing in entire institutes. For example, in 2011 the vaccine-development programme of the Sabin Vaccine Institute moved from Washington DC to co-locate at Baylor and Texas Children's Hospital.

Access to funding has given Houston

## AVENUES BEYOND ACADEMIA

### *Houston students primed for a broad range of careers*

At the University of Texas Health Science Center at Houston (UT Health), incoming graduate students in the biomedical sciences face a week-long 'boot camp'. Students are introduced to one another and to potential labmates and advisers, and are encouraged to set individual and career goals that can be used to match them with the best teams, says UT Health dean Michelle Barton.

Boot camp is the first segment of a two-part programme that is designed to

prepare students for a broad range of career options. The second, 'core' segment, called the Foundation of Biomedical Sciences, is a 15-week seminar series that includes a literature review, laboratory workshops and a discussion of research findings. Barton and joint dean Michael Blackburn developed the programme two years ago after hearing a "tidal wave of concern" about future career paths from incoming students, only a small percentage of whom planned to stay in academia. **P.S.**

institutions an impressive track record in bringing world-class institutions to town. When Rice provost George McLendon went shopping for a computational centre that could simulate the inner workings of a cell, he identified the US National Science Foundation-funded Center for Theoretical Biological Physics at the University of California, San Diego. The centre moved to Rice in 2011 after being offered \$10 million in CPRIT funding.

### TALENT SCOUTS

The move sent a message: Texas will spend money to secure talent, and the \$3-billion CPRIT endowment means that this is not just talk. "It is a huge difference to say to a top person that our state is interested in jump-starting careers like yours in this area," says McLendon. That message has made it easier for Rice to recruit 30 or so faculty members a year, half of whom are in science or engineering. Over the next five years, Rice will seek early-career scientists in bioengineering, computational sciences and physics.

Other Houston research institutions have had equal success in attracting talent. One of the latest recruits is David Sugarbaker, a cancer

researcher and mesothelioma specialist from Harvard Medical School and Brigham and Women's Hospital in Boston, Massachusetts, who arrived at Baylor in March to build a comprehensive lung-disease centre. Between now and 2019 Baylor plans to recruit 150 faculty members, says Paul Klotman, Baylor president and chief executive.

The existence of CPRIT makes recruiting easiest in cancer-related fields. But if someone identifies a big fish in a field unrelated to cancer, other state money can be tapped. Jan-Åke Gustafsson, one of the world's foremost hormone researchers, was recruited in 2009 from Stockholm's Karolinska Institute with a \$5.5-million grant to the University of Houston through the Texas Emerging Technology Fund. His institute, the Center for Nuclear Receptors and Cell Signaling (CNRCS), is jointly supported by the university and Houston Methodist Hospital.

Moving the CNRCS meant hiring more staff, says Mary Ann Ottinger, associate vice-chancellor for research at the University of Houston — and in addition, in the next several years, the university plans to recruit 30 faculty members annually across all departments, buoyed by CPRIT and similar state funds.

Houston is also attracting and launching biotechnology companies using funds from CPRIT and others, says Robert Harvey, president of the Greater Houston Partnership, the city's economic-development arm. The organization helped the London-based biotechnology company Cell Medica to open a Houston office, partly with the help of CPRIT money, Harvey says. And even if a company is not directly linked to cancer treatment, the partnership can access other incentives to help close the deal, such as the Texas Entrepreneurial Fund.

Whether Houston buys life-science entrepreneurship or builds it, the momentum is growing. Like most things in Texas, it's likely to get bigger. ■

**Paul Smaglik** is assistant editor of *Nature Careers*.



**Biomedical science is booming in Houston.**

WILADIR BULGAR/SPL/ORBIS

# AWAY

*A message to the stars.*

BY DEBBIE URBANSKI

**H**i Mom. So this is kind of weird! It's kind of like talking to myself, only in a closet! I'm joking, they make it nicer than a closet, there's a great chair, and a mirror so I can look in the mirror when I talk and pretend I'm you. The audio guys here are really great. They made me a free cap (soy) because of the wait as security has to iris scan now after last week's mob scene. Or else I bet the whole country would be here too! (Well, I know, the whole country can actually talk to you, they just have to go to their local mall booth and practically sell their kidney for some time.)

What do all those strangers say to you? In episode 2, remember, you were sitting starboard listening to messages and there was one from Idaho that made you blush? You sure have some fans down here! Dad says it's because of your outfits. He threw a fit because marriage doesn't end when you leave the planet blah blah (in the previews for next season, there is a wedding on ship!?! Will it be you?! I won't tell Dad). I have to watch what I say though because they could use this convo on the show or wherever if they feel like it. "I'll be honest. We probably won't feel like it," the audio guy said, giving me this weird look, but he made me sign the paperwork anyway. It wouldn't be fair to viewers if you had secret messages right?

Anyhow school is good, the sun came out. That was my day! Can't wait to see you tonight!

Hi. It looked like you were having this really nice time up there last night. Or is that just the editing? Do you talk about me lots and they leave those parts out? I looked at the show stats and there are mostly male viewers so maybe they think mother/daughter sappy talk (I love you I love you) would bore everybody to death (I love you). I wouldn't mind though, hint hint! I think it's gruesome how the cameras are programmed to run until the end, when everybody's, you know, dead, or even after, but I bet the producers are thinking big ratings for that. Maybe they can make it tasteful through special effects. (Sorry to be gruesome!)

I turned on every picture of you in the house. They are so old school and funny! Every time I enter a room you say: "Turn off the lights, pitch in!" Then you give a little wave like you really can see me.



At school, do you remember Larry Fields? He pushed a mike in my face and asked how does it feel to be left behind. I told him what the heck, I wasn't left. He said actually I was. He said I'm the definition of being left then he zoomed in close on my face. He asked other questions and I told him my Mom isn't really responding to me so I have no idea. He called that angle "not very interesting" so the segment probably won't run on student news tonight. It would have been fun to be on TV, too. Bye!

Hey Mom. Haven't heard from you. Are you sending me lots of messages only they get erased? Dad says you might be busy with the songs they want you to learn.

Last night we watched your application as the show

wasn't on. Okay, so we watched it ten times through. Do you still believe everything you said? The world is doomed blah blah all my potential going to waste blah let's give people some hope and find alternatives up

there blah blah. If that's true about the world, why did you leave me here? People don't think like that anymore, by the way. I mean we're all still around, right? There are still a lot of animals, at least a lot of squirrels, for sure. Dad changed all of our light bulbs over to solar too. The new thing is being hopeful.

Do you feel small up there? I'm kind of jealous of my voice, which gets to chase after you through space.

Hello again! I asked the guy who does audio is my Mom getting this? He wiggled his eyebrows and said: "Oh sure she is." Who knows what that means. The guy thinks it's sweet I come here every day. He asked am I going to do this my entire life and I said yes. They better not go out of business, especially as ratings aren't so hot. Dad took the batteries out of your pictures. He said your voice was going funny. Maybe it's better to buy nicer talkie frames next time? At least we can afford it now.

Did you know you weren't coming back when you applied? I'm guessing you forgot to read over the details, there was a lot of paper-

work, and when you got on the ship in that fancy gold outfit, doing your special kick-and-wave, the one we practised — it looked good, Mom — maybe you didn't realize you couldn't get back to us. I guess you are living your dream now, being famous etc.

I wonder if you will meet anything out there. Is there a bonus if you do? Or if you find a planet to land on? Dad won't show me the contract. I guess that's all for now. I miss you. Maybe someone will develop faster ships so in a year or two I can get on one and catch up with you. I wouldn't want to come back either then. Oops, 8.05, time to tune in. ■

**Debbie Urbanski** is a writer living in Syracuse, New York. Her stories have appeared in *Arc* (from *New Scientist*), *Interzone*, *The Kenyon Review* and the *New England Review*. [debbieurbanski.com](http://debbieurbanski.com).

JACEY

➤ **NATURE.COM**

Follow Futures:

🐦 @NatureFutures

📘 [go.nature.com/mtoodm](http://go.nature.com/mtoodm)